Determinants of Red-light Camera Violation Behavior: Evidence from Chicago, Illinois

Amr Elfar
PhD Candidate, Civil and Environmental Engineering
Northwestern University, Transportation Center
aelfar@u.northwestern.edu

Hani S. Mahmassani
(Corresponding Author)
William A. Patterson Distinguished Chair in Transportation
Director, Transportation Center
Northwestern University Transportation Center
600 Foster Street, Evanston, IL 60208, USA
Phone: +1 (847) 467-2276
Fax: +1 (847) 491-3090
masmah@northwestern.edu

Archak Mittal
PhD Candidate, Civil and Environmental Engineering
Northwestern University, Transportation Center
archakmittal2014@u.northwestern.edu

Marija Ostojic
PhD Candidate, Civil and Environmental Engineering
Northwestern University, Transportation Center
marijaostojic2015@u.northwestern.edu

Ömer Verbas
Postdoctoral Fellow
Northwestern University, Transportation Center
omer@northwestern.edu

Joseph Schofer
Professor of Civil and Environmental Engineering
Associate Dean for Academic Affairs
McCormick School of Engineering and Applied Science
Northwestern University
j-schofer@northwestern.edu

August 1, 2016

Total words: 5,157 words + 4 Tables + 2 Figures

Submitted for presentation at the 96th Annual Meeting of the Transportation Research Board, and publication in Transportation Research Record

Determinants of Red-light Camera Violation Behavior: Evidence from Chicago, Illinois

ABSTRACT

Red Light Camera (RLC) enforcement is designed to increase road safety by reducing traffic violations and crashes at road intersections. To understand the effect of traffic features, intersection factors, and signal configuration on the frequency of RLC violations, this study uses regression models to analyze violations at 152 RLCs in the city of Chicago, Illinois over a 6-year period between 2010 and 2015. The main contribution of this study is introducing panel-data analysis to better understand RLC violation behavior over time using two types of correlations in the panels (i.e. serial and spatial) that were tested to be significant in the RLC violations data. Results showed that among the factors that have a positive effect (increase) on the frequency of RLC violations are traffic volume, number of lanes, and speed limit of the approaching traffic (in direction of movement), in addition to signal cycle and an all-red phase duration of 2 seconds compared to 1. On the other hand, among the factors that have a negative effect (decrease) on the frequency of RLC violations are left-turn bays and right-on-red prohibition, in addition to a yellow-phase duration of 4 seconds compared to 3. Results also show a monthly trend in the frequency of violations where frequency is highest in Summer and lowest in Winter, and an annual learning curve where violations decrease continuously from 2010 to 2015. This paper helps decision makers and researchers in understanding the effect of different elements on violation behavior in the presence of red-light cameras.

Keywords: Red-light cameras, violations, behavior, panel data, Chicago.

INTRODUCTION

Red Light Camera (RLC) Enforcement is designed to increase vehicular safety by reducing crashes at intersections, specifically angle crashes because of their severity. In their analysis of RLC programs in the US, McFadden and McGee found that automated enforcement using RLC can result in a 20 to 60 percent reduction in traffic violations (*1*). However, RLC deployment has been the focus of considerable controversy and negative public opinion.

RLC-related studies focusing on the safety aspect are numerous in the literature, mostly in the form of "before-after" analyses where the researchers analyze the effect of RLC deployment on the number of intersection-related crashes. [See works by Lord (*2*), Walden (*3*), Washington and Shin (*4*), Hu et al.(*5*), and Retting et al. (*6*).] However, much less focus has been given by researchers towards the impact of RLCs on violation behavior.

This paper aims to contribute to the body of work on violation behavior at intersections subject to red-light camera enforcement by analyzing the different effects of traffic features, intersection factors, and signal configuration on frequencies of RLC violations at 152 camera-equipped intersections in the city of Chicago. As RLC violations were recorded over a 6-year period, panel data analysis was used to model violation frequency. Two types of correlations were assumed and tested in the models: serial (temporal) and spatial correlation. Serial correlation was considered assuming that some of the unobserved variables that affect violation behavior are correlated over time. On the other hand, spatial correlation was considered assuming that some unobserved factors affecting violation behavior could be correlated for RLCs in the same area or neighborhood.

The remaining of the paper is organized as follows. The next section provides a brief review of the different approaches used to model RLC violations in literature, followed by a description of the analyzed data set and the variables used in the regression models. Following that, the methodology is introduced for the regression models used in the analysis. Afterwards, the estimated models were discussed. Finally, the last section concludes the paper.

BACKGROUND REVIEW

Attempting to understand the reasons for RLC violations has proven to be challenging since it involves a combination of various behavioral, demographic and intersection characteristics. In general, RLC violations and crashes are negatively associated with amber light duration and width of the intersection, while positively associated with approaching flow rates and speeds (*7*). In some instances, all-red (clearance) intervals and amber phase extensions are supplementary to RLC enforcement in reducing red light violations. This practice has shown promising results according to a number of studies(*8*), (*9*), (*10*). Bonneson and Zimmerman (*10*) found that an additional 0.5 to 1.5 seconds of the amber indication interval (as long as the total time did not exceed 5.5 seconds) decreased RLC violations by up to 50%. Different models have been introduced in the literature to predict the frequency of RLC violations.

Bonneson et al. (*11*) developed a prediction model of RLC violations based on the probability distribution relative to the driver's stop or go decision which combined "exposure and contributory" factors. The model accounted for the differences among drivers due to these factors. The exposure variables were approach flow rate, number of signal cycles, and phase termination by max-out, while the contributory ones were probability of stopping and amber interval duration. The assumption was that each driver decides to go (or stop) independently of other drivers.

Hill and Lindly (*12*) tested various statistical models (linear, curvilinear, and multiple linear) to predict RLC violation frequency. Average daily traffic (ADT), number of approach lanes and speed limit were identified as the most relevant explanatory variables. However, the signal control and timing element was excluded from the analysis. Lum and Wong (*13*) applied a generalized linear model relating three independent categorical—variables, approach, lane, and time of day—to the after-red times (time-into-red), which acted as the dependent variable for the before-and-after study. Around a 40% decrease in the number of violations was observed for camera approaches; non-camera ones experienced an increase. The aggregated net reduction for all approaches was around 7%. The presence or absence of RLC significantly influenced the violation onset times (i.e. time into red) and lower mean times into red were observed for camera approaches.

Bonneson and Zimmerman (*14*), building on their previous research, examined the relationship between violation frequency and amber interval duration, indicating a trend toward more violations with shorter amber times. The authors observed the number of violations decreased with an increase in cycle length, amber indication duration, volume-to-capacity (V/C) ratio, intersection width, speed etc. Most interestingly, the authors found the lowest number of violations were associated with V/C ratios in the range of 0.6 to 0.7, regardless of any other significant factor value.

Yang and Wassim (*15*) built a logistic regression model in order to understand the relation between red light violations and various driver, intersection, and environmental factors. They reported that approximately 56 % of the violators traveled at or below the posted speed limit. Additionally, violations occurred 94 % of the time within 2 seconds after the onset of the red light. The authors' findings confirmed older drivers were more likely to run a red light than younger drivers when the elapsed time since the onset of red light was more than 2 seconds.

The most recent approach in RLC violation prediction studies involves using observational data supplemented with driving simulator data. Jahangiri et al. (*16*) adopted a random forest (RF) machine-learning technique to develop RLC violation prediction models. The majority of the previous research efforts, however, recognized the limitations of the models suggested. This was predominately related to the types of models and variables used and "local" prediction model calibration issues (that is, models not robust enough to be transferable to other areas and/or geometry configurations).

DATA

The Chicago Department of Transportation provided the data for this study. Information related to 152 RLCs at 85 four-legged intersections were retrieved. Locations of the RLC intersections are shown in FIGURE 1. Time period covered range between 2010 and 2015. In this date range, all of the violations were provided with date-time stamp for all the cameras, except for maintenance and black-out periods where violations were not detected. The dataset included: date-time, speed of the vehicle while violating, associated vehicular lane and posted speed limit. Information related to signal timing contains the all red duration, yellow time, cycle length, total number of lanes on the approach.

Necessary additional information was readily available through online resources. Google Maps was used to manually obtain intersection geometry and configuration related information. These included intersection traverse distance, type of median, presence of dedicated left turn arrow, right turn on red prohibition sign, left and right turn bays. Annual Average Daily Traffic

(AADT) was obtained from an online data portal provided by the city of Chicago; however, we corrected AADT for monthly traffic patterns as published by the Illinois Department of Transportation for the different years.

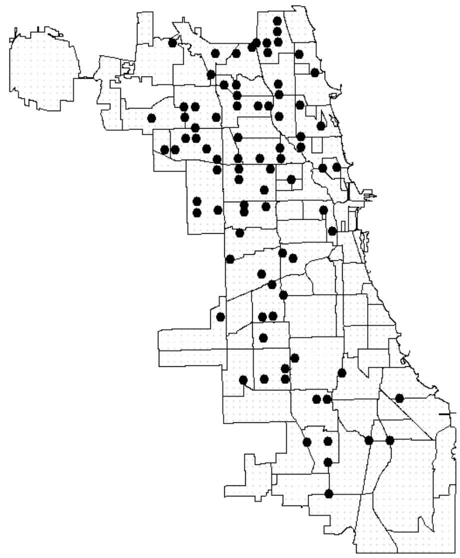

FIGURE 1 LOCATIONS OF RLC INTERSECTIONS IN CHICAGO, IL

Variables in Regression Models

In the regression model, the dependent variable was defined as the number of RLC violations per month, where N = 152 cameras (panels) and T = 72 time periods (months). To test for different RLC violation behaviors, four classes of violation were defined: All violations, Right-On-Red (ROR), High speed, and One-sec-into-red. All violations include all observed RLC violations for on an approach by a specific camera. ROR violations includes cases where a vehicle turned right when "NO TURN ON RED" sign is present while signal is red. High-speed violation includes cases where a vehicle run an RLC with speed that is more than 10 percent above speed limit. One-sec-into-red includes cases where a vehicle run an RLC within 1 sec after the signal had turned red.

Table 1 presents a summary of the variables included in the regression model. Three directions of movement were defined for the variables relative to the movement of a vehicle approaching an RLC: self, crossing and opposite. Self indicates that the variable, for example speed limit, describes the approach on which the vehicle is moving towards an intersection. Crossing describes the approach that is crossing (perpendicular to) the self-approach on an intersection. Oppsite describes the approach that is opposite of the self-approach.

Missing Data

The data set includes 10,944 observations (152 x 72), for 152 red-light cameras (panels) over 72 months. Due to maintenance and short black-out periods of some cameras, violations were not detected for specific time periods. As model specifications of spatial and serial correlations require a balanced panel data set where the same number of time periods is available for all panels, a multiple imputations algorithm was implemented to fill in missing observations of RLC violations based on the trends of the known observations. Although missing observations account for only 3.4% percent of the total observations in the data set, using a multiple imputations should reduce the bias that might result from missing observations or using a simple average to fill them (*17*). One concern was that the imputed values were of the dependent variable rather than explanatory variables of which no data was missing. However, as Young, Johnson, and Graham (*18*) (*19*) explain, an imputation model does not capture causal relationships in the data. Rather a tool to "preserve important features of observed information in imputed values" (*18*).

TABLE 1 Description of Variables in Regression Models

Variable	Description	Mean	S.D
Dependent Variable			
All vio.	Continuous: All RLC violations per month time period	129.66	118.28
ROR	Continuous: ROR violations per month time period	16.17	40.89
High-speed Violations	Continuous: High-speed violations per month time period	27.18	48.56
1-into-red violations	Continuous: One-sec-into-red violations per month time period	63.19	67.04
Explanatory Variables			
AADT/lane - self	Continuous: Average Annual Daily Traffic per lane, corrected for monthly traffic patterns in the (self) direction	6.37	2.13
AADT/lane - crossing	Continuous: Average Annual Daily Traffic per lane, corrected for monthly traffic patterns in the (crossing) direction	6.16	2.27
N. lanes - self	Continuous: Number of lanes in (self) direction	3.25	1.01
N. lanes - crossing	Continuous: Number of lanes in (crossing) direction	3.12	1.04
Speed limit - self	Continuous: Speed limit in (self) direction	30.53	2.64
Speed limit - crossing	Continuous: Speed limit in (crossing) direction	30.46	2.83
Traverse Distance - self	Continuous: Intersection traverse distance in (self) direction	99.00	19.95
Traverse Distance - crossing	Continuous: Intersection traverse distance in (crossing) direction	101.63	19.54
Left-turn bay – self	Binary: Indicator of existing left-turn bay in (self) direction	0.90	0.30
Left-turn blocked	Binary: Indicator of prohibited left turn movement in (self) direction	0.02	0.14
Left-turn arrow – oppst.	Binary: Indicator of existing left turn arrow for opposite approach	0.57	0.50
ROR prohibition - self	Binary: Indicator of existing "NO TURN ON RED" sign	0.48	0.50
Right-turn bay - self	Binary: Indicator of existing right-turn bay in (self) direction	0.08	0.27
Median - self	Binary: Indicator of existing median (physical or yellow line)	0.22	0.41
Cycle length	Continuous: Length of signal cycle in seconds	86.67	16.83
Yellow phase	Factor: Length of yellow phase in seconds (3 or 4 sec)	3.05	0.21
All-red phase	Factor: Duration of all-red phase in seconds (1 or 2 sec)	1.65	0.48
Month	Factor: Indicator of the month for the time period (1 -12)	-	-
Year	Factor Indicator of the year for the time period (2010 – 2015)	-	-

The implemented algorithm, AMELIA (a package in R system), performs multiple imputations for each missing cell in the data set based on observed data to create a complete data set. The multiple imputations capture the uncertainty in the missing data. AMELIA has two main assumptions behind its algorithm: 1) complete data are multivariate normal, 2) data are missing at random (MAR). MAR means that the "pattern of missingness depends on the observed data"(*17*). Thirty imputations were performed for each missing cell, and the average of those 30 imputations was used to fill the missing data. The creators of the algorithm suggest that 5 imputations are enough for most data sets, however, 30 imputations were used to reduce uncertainty. More information on the imputation algorithm can be found in (*17*).

METHODOLOGY

To model frequency of RLC violations in Chicago IL, two types of regression models were used: serially correlated (time-dependent) panels, and spatially correlated panels. Panel data analysis (often referred to as longitudinal or cross-sectional time series data) was chosen since RLC violations were observed over a significant period of time (6 years). This section explains the specification of the models used in our analysis.

Serially Correlated Panels

The assumption behind serial correlation is that some unobserved factors that affect violation behavior are correlated over time. To capture that, a first-order serial autocorrelation parameter was specified in the error term of a pooled linear regression model (*20*). Individual (fixed) effects model was disregarded since all RLCs are located in Chicago, IL and are setup at comparable signalized intersections. The model specification is as follows:

$$y_{i,t} = x_{i,t}\beta + \nu_{i,t} \quad (1)$$

$$\nu_{i,t} = \rho_i \nu_{t-1} + e_{i,t} \quad (2)$$

where i = 1,…,N cameras, t =1,…,T time-periods, $y_{i,t}$ is the frequency of RLC violations for camera i and time-period t, $x_{i,t}$ is a vector of explanatory variables (AADT, road geometry, and signal timing variables) with coefficients β, ν is a vector of first-order serially autoregressive errors (AR1) with ρ_i as the serially autoregressive parameter for camera i.

Spatially Correlated Panels

Since RLC violations were recorded for cameras in different areas of Chicago, one can assume that some unobserved factors that affect frequency of violations are correlated for cameras that are in the same neighborhood or area. One way to capture the spatial interaction is introducing a spatially structured autocorrelation parameter to the error term in an ordinary panel regression (*21*). As cameras are all in Chicago with very similar characteristic, panel specific effect were ignored. The model general formula is as follows:

$$y = X\beta + \varepsilon \quad (3)$$

$$\varepsilon = \rho(I_T \otimes W_N)\varepsilon + e \quad (4)$$

where y is an $NT \times 1$ vector of observations on the dependent variable (RLC violation per month), X is a $NT \times k$ matrix of observations exogenous explanatory variables (traffic features, intersection factors, and signal configuration), I_T is an identity matrix of dimension T, W_N is the $N \times N$ spatial weights matrix. ε is a vector of spatially autoregressive errors that follow a spatial autoregressive process of the form described in formula (4) with ρ as the spatial autoregressive parameter, W_N the spatial weights matrix and $e \sim IID(0, \sigma_e^2)$, and $\otimes$ is the Kronecker product, an operation on two matrices of arbitrary sizes.

The spatial weight matrix was created using Euclidian distances between cameras. The inverse of square root distance was used to create spatial correlation between cameras, normalized by the total inverse distances to have correlation be between 0 and 1. The assumed structure indicates that spatial correlation decreases as distance increases. Different weight structures were tested, however, the inverse squared root distance resulted in significant spatially autocorrelated parameter.

MODEL ESTIMATES AND RESULTS

Linear Regression with Serial Correlation

Generalized Least Squares (GLS), built in the statistical software STATA, was used to estimate the total RLC violations models. GLS performs better at estimating effects in time-series data when heteroscedasticity and serial correlation are significant (*20*).

Testing for heteroscedasticity and serial correlation

The log-likelihood ratio (LR) test was used to test for significance of heteroscedasticity. To do so, two models were estimated: one assuming heteroscedastic panels and another assuming homoscedastic panels. To estimate log-likelihoods of the models, the iterated GLS option was used in STATA where maximum-likelihood estimates are produced. The LR chi-squared value for the test was 10845.33 with p-value equal 0.00 for 151 degrees of freedom at the 0.05% significance level, indicating significant heteroscedasticity in the data.

For testing serial correlation, Wooldridge's test of autocorrelation in panel data was used (*22*). Wooldridge uses the F statistic to test the null hypothesis that no first-order autocorrelation exists in the data. The F statistic value for the total violations model was 484.24 with p-value equal 0.00 for (1,151) degrees of freedom at 0.05% significance level, indicating significant serial correlation.

Model Estimates

Table 2 shows the estimated model for total violations, with significant variables retained at the 0.05% level. The Wald chi-squared statistic is 1993.39 with 0.00 probability being larger than critical chi-squared for 32 degrees of freedom, indicating an overall significant model. Self, crossing, and oppst in the variable names indicate the direction of car/traffic movement at an intersection as explained in section 3.

The model shows that variables which have a positive effect (increase) on the frequency of RLC violations are AADT/lane self, N. lanes self, speed limit, traverse distance-crossing, blocked left turn, cycle length, and all-red phase of 2 sec compared to 1 sec. On the other hand, variables which have a negative effect (decrease) on the frequency of RLC violations include AADT/lane - crossing, N. lanes - crossing, traverse distance – self, left-turn bay left-turn arrow – oppst, ROR-prohibition, median, and a yellow phase of 4 seconds compared to 3. Furthermore, the model shows a monthly trend in the frequency of violations where frequency is highest in Summer and lowest in Winter, and an annual learning curve where violations decrease continuously from 2010 to 2015

TABLE 2 MODEL ESTIMATE FOR ALL VIOLATIONS ASSUMING SERIAL CORRELATION

Variable	Coefficient	Std. Error	z	P>|z|	[95% Conf. Interval]	
AADT/lane - self	1.72	0.64	2.68	0.01	0.46	2.97
AADT/lane - crossing	-1.71	0.67	-2.57	0.01	-3.02	-0.41
N. lanes - self	26.13	2.13	12.24	0.00	21.94	30.31
N. lanes - crossing	-6.25	2.30	-2.71	0.01	-10.76	-1.73
Speed limit - self	4.04	0.58	7.03	0.00	2.91	5.17
Speed limit - crossing	2.85	0.62	4.63	0.00	1.64	4.06
Traverse Distance - self	-0.63	0.13	-5.00	0.00	-0.87	-0.38
Traverse Distance - crossing	0.95	0.13	7.24	0.00	0.69	1.21
Left-turn bay - self	-24.06	5.88	-4.09	0.00	-35.58	-12.53
Left-turn blocked - self	61.37	12.29	4.99	0.00	37.28	85.46
Left-turn arrow – oppst.	-30.94	3.94	-7.86	0.00	-38.65	-23.22
ROR prohibition - self	-24.51	3.20	-7.66	0.00	-30.78	-18.24
Median - self	-15.91	5.47	-2.91	0.00	-26.63	-5.20
Cycle length	1.26	0.16	8.01	0.00	0.95	1.57
Yellow phase =4	0			(Reference)		
Yellow phase =4	-108.80	7.83	-13.89	0.00	-124.15	-93.45
All-red phase =1	0			(Reference)		
All-red phase =2	10.11	3.97	2.54	0.01	2.32	17.90
Month						
1	0			(Reference)		
2	-0.91	0.86	-1.07	0.29	-2.59	0.76
3	11.44	1.30	8.82	0.00	8.90	13.99
4	13.61	1.51	9.01	0.00	10.65	16.57
5	26.42	1.58	16.76	0.00	23.33	29.51
6	27.98	1.66	16.82	0.00	24.72	31.24
7	29.93	1.72	17.40	0.00	26.56	33.30
8	26.20	1.63	16.06	0.00	23.00	29.40
9	18.41	1.53	12.04	0.00	15.41	21.41
10	12.79	1.50	8.52	0.00	9.85	15.73
11	4.51	1.39	3.25	0.00	1.79	7.23
12	1.03	1.08	0.95	0.34	-1.08	3.14
Year						
2010	0			(Reference)		
2011	-11.96	1.96	-6.09	0.00	-15.81	-8.11
2012	-22.42	2.53	-8.87	0.00	-27.37	-17.47
2013	-28.80	2.84	-10.13	0.00	-34.38	-23.23
2014	-34.25	3.22	-10.62	0.00	-40.57	-27.93
2015	-34.66	3.63	-9.55	0.00	-41.77	-27.55
Intercept	-237.71	33.42	-7.11	0.00	-303.20	-172.21

Starting with traffic features, AADT/lane – self and N. lanes – self can be interpreted as exposure variables whose positive coefficients (1.72 and 26.13 respectively) indicate that higher traffic leads to higher chances of RLC violations. The negative coefficients of AADT/lane – crossing and N. lanes – crossing (-1.71 and -6.25 respectively) indicate that it might be harder for drivers to violate when crossing traffic is high. The positive coefficients of Speed limit – self/crossing (4.04 and 2.85) shows that at higher speeds, drivers can be more confident to cross an intersection in time with the risk of a violation.

As for intersection factors, the negative coefficient of Traverse distance -self (-0.63) indicates that a longer distance to traverse an intersection makes it harder for drivers to pass through an intersection in time, hence, less likely to violate. On the other hand, the positive coefficient of Traverse distance -crossing (0.95) indicates that a wider intersection would make drivers more confident to pass through it before crossing traffic starts moving increasing the chances of a violation. Left-turn bay - self has a negative coefficient (-24.06) indicating that drivers are less likely to violate RLC after impatiently waiting behind a vehicle turning left if turn bay exist. Left-turn blocked - self has a positive coefficient (-24.06) which could mean that drivers are more confident in passing through an intersection, risking a violation, without worrying about crossing traffic from one direction. Left-turn arrow – oppst has a negative coefficient (-30.94) indicating that drivers are less likely to violate, and risk a crash, when the number of left-turning vehicles are high in the opposite direction. This is under the assumption that a left-turn arrow is installed when the number of turning vehicles is high. Right-On-Red prohibition and median have negative coefficients (-24.51 and -15.91 respectively) indicating that when installed, violation frequency decreases.

Regarding the effect of signal configuration, the positive coefficient of cycle length (1.26) shows that higher cycle length could make people impatient and more likely to violate a RLC. The negative yellow phase coefficient (-108.80) shows that increasing the phase length to 4 seconds instead of 3 reduces the number of violations. Longer yellow phase duration increases the probability of drivers passing through an intersection before signal turns red, avoiding a violation. All-red phase, while being important for safety, can be interpreted as an exposure variable whose positive coefficient (10.11) indicates that increasing all-red duration from 1 to 2 seconds increases the probability that a violation occurs.

Predicted vs. actual values of total RLC violations are plotted in FIGURE 2 for the 72-month time periods using the serially correlated model. The plot shows that the model (black bars) picks up the annual and monthly trends in RLC violations, however, it tends to flatten out the spikes in numbers as expected of a linear regression model. It is worth noting that the annual and monthly trends of actual violation numbers are consistent and decreasing over the years.

In addition to the total RLC violations model, separate models were estimated for three classifications (defined in section 3) of RLC violations: right-on-red, high speed, and one-sec-into-red. The separate models were estimated to test whether explanatory variables have different effects on the different classification of violations. TABLE 3 summarizes the significant coefficients at the 0.05% level for the four different models. Generally, the violation behavior is similar for the different classes of violations in terms of effect sign (increasing/decreasing) despite different magnitudes. The different coefficient magnitudes capture the different quantities of violation classes, but should not affect the direction of the effect (positive/negative). Furthermore, some variables were insignificant for specific classes while significant for others. It is worth noting that the classes of violations are not mutually exclusive nor exhaustive. For example, a high speed violation can be a 1-sec-into-red as well. Additionally, some violations were not classified into any of the 3 classification defined earlier, but are included in the all violations model.

TABLE 3 Model Estimates for RLC violation Classes using serial Correlation

Variable	**Estimated Models**			
	All vio.	**ROR**	**High-speed**	**1-into-red**
AADT/lane - self	1.7	-1	1.2	2.9
AADT/lane - crossing	-1.7	0.8	-1	-1
N. lanes - self	26.1	9.8	7.1	16.2
N. lanes - crossing	-6.2	5.9	-4.5	-6.5
Speed limit - self	4	1.7	0	1.8
Speed limit - crossing	2.9	0	0	0
Traverse Distance - self	-0.6	0	-0.1	-0.3
Traverse Distance - crossing	1	0	0.4	0.6
Left-turn bay – self	-24.1	0	0	0
Left-turn blocked - self	61.4	0	0	44.5
Left-turn arrow - oppst	-30.9	0	-3.2	-11.8
ROR prohibition - self	-24.5	0	-4.8	-6.4
Median - self	-15.9	-9.5	0	-7.1
Cycle length	1.3	0.6	-0.07	0.3
Yellow phase =3	0		(Reference)	
Yellow phase =4	-108.8	-20.1	-24	-56.3
All-red phase =1	0		(Reference)	
All-red phase =2	10.1		6.1	5.3
Month				
1	0		(Reference)	
2	-0.9	-1	-0.7	-0.4
3	11.4	2.4	0.6	3.8
4	13.6	3.3	0.5	4.6
5	26.4	7	1.5	10.4
6	28	6.9	1.5	11.2
7	29.9	7.3	2.2	11.9
8	26.2	5.2	1.8	11.5
9	18.4	3.9	0.7	9
10	12.8	3	0.1	6.3
11	4.5	1.3	-0.6	7.3
12	1	0.5	-1.1	6.3
Year				
2010	0		(Reference)	
2011	-12	-2.4	-1.6	0.4
2012	-22.4	-4.3	-4.1	2
2013	-28.8	-5.6	-6.8	6.9
2014	-34.3	-8.3	-10.1	17.4
2015	-34.7	-9	-12.2	23.8
Right-turn bay - self		20	7.1	9.7
Intercept	-237.7	-116.6	-5.5	-100.8

An exception to the general behavior is the effect of AADT/lane on ROR violations where higher traffic in the direction of movement (AADT/lane – self) decreases the likelihood of an ROR violation while higher crossing traffic increases the chances of an ROR violation. This could indicate more opportunities to turn right on red with higher crossing traffic. Another interesting exception to the general behavior is the annual learning curve to 1-sec-into-red violations. The model shows that 1-sec-into-red violations are increasing over the years despite the general trend an annual decrease in all violations.

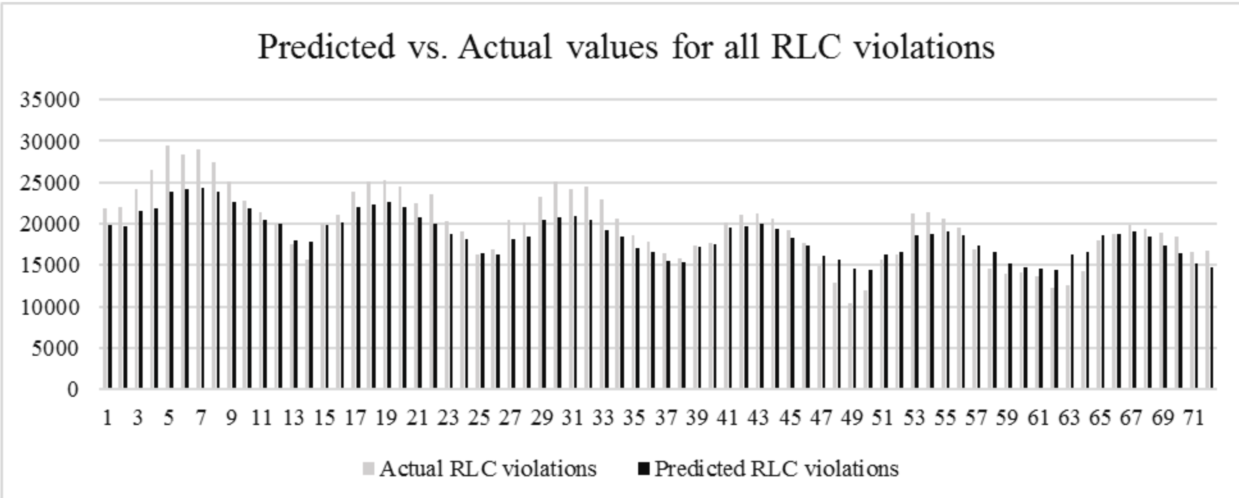

FIGURE 2 Predicted vs. Actual Values of Total RLC Violations using Serially Correlated Model

Linear Regression with Spatial Correlation

Maximum likelihood estimation (ML) was used to estimate violation regression model assuming a spatially autoregressive error term, as discussed in the methodology section. The estimator tool was developed by Giovanni Millo as a package for the statistical computing system R. Details on the likelihood functions and using the tool can be found in Millo paper (*21*).

Testing for spatial correlation

As a first step, spatial autocorrelation was tested for the assumed spatial weight structure. To that end, we applied the conditional Lagrange Multiplier (LM) test developed by Beltagi et al. (*23*) and built in Millo's R package (*21*). The conditional LM tests the null hypothesis that the spatial autocorrelation coefficient is zero assuming the random effects may or may not be present. The alternative hypothesis is that the spatial autocorrelation coefficient does not equal zero. The LM_{λ} statistic value was 91.978 with p-value equal 0.00, in which case the spatial autocorrelation is significantly different from zero.

Model Estimate

The ML estimate model spatially correlated panels (ML-SP) is shown in TABLE 4, along with the GLS estimate for serially correlated panels and an Ordinary Least Squares (OLS) estimate assuming no correlation. To test for the overall ML-SP model significance against the null model, the log-likelihood ratio test was used. The LR chi-squared value was 3003.5 with p-value equal 0.00 for 32 degrees of freedom at the 0.05% level, indicating that the model is overall significant.

The ML-SP estimate shows that, while spatial autocorrelation parameter (rho) is significantly different from zero at the 0.05% level (0.08), the effect on the estimated coefficients is negligible compared to OLS (no correlation). On the contrary, the effect of specifying a serially autocorrelated error term is a significant change in coefficient estimates compared to OLS, indicating that serial correlation is much more dominating than spatial correlation in the data set. It is worth noting that different spatial weight structures were tested, some of which yielding rho values that are significantly different from zero. However, the effect on coefficients was also negligible.

TABLE 4 Comparison of Spatially, Serially, and Non-correlated models

Variable		ML-SP	GLS-SR	OLS
Rho		0.08	0.83	-
AADT/lane - self		4.00	1.72	3.96
AADT/lane - crossing		-4.20	-1.71	-4.19
N. lanes - self		26.66	26.13	26.53
N. lanes - crossing		-8.43	-6.25	-8.43
Speed limit - self		4.98	4.04	4.94
Speed limit - crossing		2.39	2.85	2.43
Traverse Distance - self		-0.57	-0.63	-0.57
Traverse Distance - crossing		1.35	0.95	1.35
Left-turn bay - self		-19.61	-24.06	-19.53
Left-turn blocked - self		37.78	61.37	37.19
Left-turn arrow – oppst.		-31.08	-30.94	-31.68
ROR prohibition - self		-16.49	-24.51	-16.30
Median - self		-27.83	-15.91	-28.24
Cycle length		1.57	1.26	1.59
Yellow phase =4		0.00	(Reference)	
Yellow phase =4		-120.22	-108.8	-119.34
All-red phase =1		0.00	(Reference)	
All-red phase =2		4.74	10.11	4.76
Month				
	1	0.00	(Reference)	
	2	-2.08	-0.91	-2.07
	3	15.29	11.44	15.30
	4	21.54	13.61	21.56
	5	43.40	26.42	43.41
	6	47.91	27.98	47.93
	7	48.02	29.93	48.03
	8	43.48	26.2	43.48
	9	32.02	18.41	32.04
	10	23.59	12.79	23.64
	11	10.44	4.51	10.51
	12	4.70	1.03	4.75
Year				
	2010	0.00	(Reference)	
	2011	-21.62	-11.96	-21.60
	2012	-25.96	-22.42	-25.86
	2013	-45.50	-28.8	-45.36
	2014	-54.71	-34.25	-54.55
	2015	-53.34	-34.66	-53.13
Intercept		-313.78	-237.71	-312.47

CONCLUSION

While understanding the safety implications of RLC enforcement is essential, as reflected by the existing literature, another important (and overlooked) subject is understanding how different elements affect violation behavior in the presence of RLCs and how that behavior changes over time. This paper aims at answering those questions by using regression models for panel data to infer the effect of traffic features, intersection factors, and signal configuration on the frequency of Red-light Camera (RLC) violations and the change of frequency over time.

To that end, the study analyzed RLC violations at 152 cameras at 85 intersections in the city of Chicago, IL over 72-month period (2010 – 2015). Two types of regression models for panel data were introduced: serially correlated panels (time-dependent) which was estimated by the Generalized Least Squares (GLS) method, and spatially correlated panels estimated by the Maximum Likelihood Estimation (MLE) method. However, only serial correlation showed significant effect on coefficient estimates.

Results showed that variables which have a positive effect (increase) on the frequency of RLC violations are AADT/lane - self, N. lanes - self, Speed limit, Traverse distance - crossing, blocked left turn, cycle length, and all-red phase of 2 sec compared to 1 sec. On the other hand, variables which have a negative effect (decrease) on the frequency of RLC violations include AADT/lane - crossing, N. lanes - crossing, Traverse distance – self, Left-turn bay, Left-turn arrow – oppst, ROR-prohibition, median, and a yellow phase of 4 seconds compared to 3. Results also show a monthly trend in the frequency of violations, and an annual learning curve where violations decrease continuously from 2010 to 2015.

Furthermore, accounting for annual and monthly effects, models showed that RLC violations were continuously decreasing over the studied years, thus indicating a positive change in violation behavior. Additionally, monthly effects were significant, indicating other unobserved variables in the data, like weather, could affect number of RLC violations per month.

The findings of this paper help policy makers and researchers understand the interactions of different elements with RLC violation behavior. While the introduced models try to explain violation behavior in the city of Chicago, the methodology can be used to build models to explain RLC violation behavior in other areas. However, the general direction of effects (positive/negative) of the considered factors confirms results found in literature for other cities. For future work, a survey can be done to collect drivers' insights on how the significant factors found in this study affect their driving behavior at RLCs. Drivers' insights would improve the interpretation of results discussed in the study. In addition, models in this study could be extended using virtual reality tools, like driving simulators, to test for effect of unobserved elements in this study.

ACKNOWLEDGEMENT

This paper is based on research funded by the Chicago Department of Transportation (CDOT) as part of a project titled "Red Light Camera Enforcement: Best Practices and Program Road Map" conducted by the Northwestern University Transportation Center (NUTC). The authors are grateful to the cooperation of Abraham Emmanuel, Lawrence McPhillips, Yadollah Montazery, and Kevin O'Malley at the CDOT. The authors also acknowledge the contributions of Mr. Breton Johnson at the NUTC. The work has benefited from discussion with Drs. Larry Head (University of Arizona), Fred Mannering (University of South Florida), and Troy Walden (Texas Transportation Institute), who served as advisors to the NUTC researchers. The authors

naturally remain solely responsible for the content of this paper; any views and opinions are those of the authors and do not necessarily reflect those of CDOT.

REFERENCES

[1] McFadden, J., and H. W. McGee. Synthesis and evaluation of red light running automated enforcement programs in the United States.In, 1999.

[2] Lord, D. Safety Effects of the Red-Light Camera Enforcement Program in Chicago, Illinois. 2014.

[3] Walden, T. Effectiveness of Red Light Cameras-Texas Statewide Evaluation. *Institute of Transportation Engineers. ITE Journal,* Vol. 81, No. 12, 2011, p. 30.

[4] Washington, S., and K. Shin. *The impact of red light cameras (automated enforcement) on safety in Arizona*. Arizona Department of Transportation, 2005.

[5] Retting, R. A., and S. Y. Kyrychenko. Reductions in injury crashes associated with red light camera enforcement in Oxnard, California. *American journal of public health,* Vol. 92, No. 11, 2002, pp. 1822-1825.

[6] Hu, W., A. T. McCartt, and E. R. Teoh. Effects of red light camera enforcement on fatal crashes in large US cities. *Journal of safety research,* Vol. 42, No. 4, 2011, pp. 277-282.

[7] Research, I. f. S. City of Albuquerque Yellow Light Timing Change and All-Red Clearance Interval Timing Change Effectiveness Study - Final Report. *The City of Albuquerque Department of Municipal Development and the Office of the Mayor*, 2012.

[8] Retting, R. A., S. A. Ferguson, and C. M. Farmer. Reducing red light running through longer yellow signal timing and red light camera enforcement: results of a field investigation. *Accident Analysis & Prevention,* Vol. 40, No. 1, 2008, pp. 327-333.

[9] Retting, R., and M. Greene. Influence of traffic signal timing on red-light running and potential vehicle conflicts at urban intersections. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1595, 1997, pp. 1-7.

[10] Bonneson, J., and K. Zimmerman. Effect of yellow-interval timing on the frequency of red-light violations at urban intersections. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1865, 2004, pp. 20-27.

[11] Bonneson, J. A., K. Zimmerman, and M. A. Brewer. Engineering countermeasures to reduce red-light-running.In, Texas Transportation Institute, Texas A & M University System, 2002.

[12] Hill, S. E., and J. K. Lindly. Red light running prediction and analysis.In, the Center, 2003.

[13] Lum, K., and Y. Wong. Impacts of red light camera on violation characteristics. *Journal of transportation engineering,* Vol. 129, No. 6, 2003, pp. 648-656.

[14] Bonneson, J. A., and K. Zimmerman. Development of guidelines for identifying and treating locations with a red-light-running problem.In, Texas Transportation Institute, Texas A & M University System, 2004.

[15] Yang, C., and W. G. Najm. Analysis of red light violation data collected from intersections equipped with red light photo enforcement cameras.In, 2006.

[16] Jahangiri, A., H. Rakha, and T. A. Dingus. Red-light running violation prediction using observational and simulator data. *Accident Analysis & Prevention*, 2016.

[17] Honaker, J., G. King, and M. Blackwell. Amelia II: A program for missing data. *Journal of Statistical Software,* Vol. 45, No. 7, 2011, pp. 1-47.

[18] Young, R., and D. R. Johnson. 'Imputing the Missing Y's: Implications for Survey Producers and Survey Users.In *Proceedings of the AAPOR Conference Abstracts*, 2010. pp. 6242-6248.

[19] Graham, J. W. Missing data analysis: Making it work in the real world. *Annual review of psychology,* Vol. 60, 2009, pp. 549-576.
[20] Greene, W. H. Econometric Analysis. 2012.
[21] Millo, G., and G. Piras. splm: Spatial panel data models in R. *Journal of Statistical Software,* Vol. 47, No. 1, 2012, pp. 1-38.
[22] Drukker, D. M. Testing for serial correlation in linear panel-data models. *Stata Journal,* Vol. 3, No. 2, 2003, pp. 168-177.
[23] Baltagi, B. H., S. H. Song, and W. Koh. Testing panel data regression models with spatial error correlation. *Journal of econometrics,* Vol. 117, No. 1, 2003, pp. 123-150.