Article

Identifying Temporal Aggregation Effect on Crash-Frequency Modeling

Bumjoon Bae [1], Changju Lee [2,†] , Tae-Young Pak [3] and Sunghoon Lee [4,*]

1. Center for Privately-Financed Highway Studies, The Korea Transport Institute, Sejong 30147, Korea; bbae2016@koti.re.kr
2. Environment, Planning and Economic Division, Virginia Transportation Research Council, Charlottesville, VA 22904, USA; cl8ax@virginia.edu
3. Department of Consumer Science, Sungkyunkwan University, Seoul 03063, Korea; typak@skku.edu
4. Business Data Analytics Team, Samsung Card Co., Ltd., Seoul 04514, Korea
* Correspondence: sunghoonleebest@gmail.com
† Current Affiliation: Transport Division, United Nations Economic and Social Commission for Asia and the Pacific, Bangkok 10200, Thailand.

Abstract: Aggregation of spatiotemporal data can encounter potential information loss or distort attributes via individual observation, which would influence modeling results and lead to an erroneous inference, named the ecological fallacy. Therefore, deciding spatial and temporal resolution is a fundamental consideration in a spatiotemporal analysis. The modifiable temporal unit problem (MTUP) occurs when using data that is temporally aggregated. While consideration of the spatial dimension has been increasingly studied, the counterpart, a temporal unit, is rarely considered, particularly in the traffic safety modeling field. The purpose of this research is to identify the MTUP effect in crash-frequency modeling using data with various temporal scales. A sensitivity analysis framework is adopted with four negative binomial regression models and four random effect negative binomial models having yearly, quarterly, monthly, and weekly temporal units. As the different temporal unit was applied, the result of the model estimation also changed in terms of the mean and significance of the parameter estimates. Increasing temporal correlation due to using the small temporal unit can be handled with the random effect models.

Keywords: modifiable temporal unit problem (MTUP); crash-frequency modeling; traffic safety; negative binomial regression

1. Introduction

Aggregation of spatiotemporal data allows researchers to save efforts in collecting data and modeling crash-frequency attributes efficiently. However, aggregating data can encounter potential information loss or distort attributes via individual observation, which would influence modeling results and lead to an erroneous inference, named the ecological fallacy. Therefore, the decision of spatial and temporal resolution is a fundamental consideration in a spatiotemporal analysis.

For spatial aggregation, much research has been conducted under the name of the modifiable area unit problem (MAUP). The MAUP stems from a zoning system used to collect geographical data and consider modifiable areal units in the analysis [1]. Corresponding to the MAUP, the modifiable temporal unit problem (MTUP) occurs when using data that is temporally aggregated. While consideration of the spatial dimension has been increasingly studied, the counterpart, the temporal unit, is rarely considered, particularly in the traffic safety modeling field.

The spatial aggregation scale and configuration can be decided in numerous ways, and therefore modeling results can be different depending on the scale and configuration. Depending on the temporal aggregation scale, the variability of variables can be decrease or

increase. The variability increases when a variable is aggregated by a small temporal unit and vice versa. These characteristics caused by choice of temporal scale affect modeling results and could lead to an erroneous inference. For example, in crash-frequency modeling, when variables are aggregated using a one-year unit, seasonal variation, such as increases in crash-frequency in rainfall or heavy snow seasons, could not be captured compared to variables aggregated using a one-quarter unit. Therefore, crash modeling results could not also reflect the seasonal variation.

Particularly, crash-frequency models in the field of crash modeling have been considered with diverse time-varying explanatory variables. The most common variables are annual average daily traffic (AADT) and/or vehicle miles traveled (VMT) as exposure variables [2–5]. Ivan et al. [4] used light condition and v/c ratio among day and time-of-day dummy variables as temporal factors of crash-frequency models, although those models belong to a relatively disaggregate level. Socio-demographic factors, such as population, number of households, number of employees, and number of registered cars, are also time-varying, as well as traffic characteristics [6].

Using data with different temporal units will bring the different extent of information loss into the modeling process, which refers to unobserved heterogeneity and/or temporal correlation effects [7,8]. It is obvious that time-varying variables, such as traffic volume and speed, have the nature of temporal autocorrelation. However, non-time-varying variables, such as the same roadway characteristics, will produce many observations with different crash frequencies, which will also be correlated over time due to the remaining unobserved factors related to the variables [8]. To address the temporal effect properly, selecting a temporal scale should be paid careful attention to in the modeling process [9].

The issue of temporal aggregation of data called the modifiable temporal unit problem (MTUP), has been commonly neglected, though it has an impact on crash-frequency modeling. Meanwhile, there has been growing attention to the spatial counterpart, called the modifiable areal unit problem (MAUP). The MAUP occurs when the spatial zoning systems used to collect spatial data are arbitrary [1]. Miller [1] suggested three available approaches associated with the MAUP in transport demand modeling, which are assessing zoning system effects, designing optimal zoning systems, and deriving better zonal distance measures. Zhang and Kukadia [10] showed that the MAUP effect can be divided into two sub-effects: scale effect and zonal effect. The scale effect refers to using different aggregation scales, and the zonal effect is related to using different zoning configurations [10]. Xu et al. [11] conducted a sensitivity analysis using 15 spatial aggregation schemes for the study area to quantify the MAUP effect in regional crash-frequency modeling. The results show that as the number of zones increases, the spatial autocorrelation of crash data increases by using Moran's I, and the estimates of parameters are more stable in terms of statistical significance and standard error [11].

It is noted that there are a growing number of studies in various fields, such as ecology, economics, political science, and geography, that deal with the MTUP effect because it is as crucial a part of the spatiotemporal analysis as the MAUP. Jong and Bruin [12] found that aperiodicity in the data could influence model results and, therefore, indicated the temporal aggregation needs to be carried out with care to avoid spurious model results. Koch and Carson [13] claimed that, in a sparsely populated area, the scale in a temporal context was important as well as the scale of space. Helbich et al. [14] stressed that, due to the MTUP, inference to assess associations between COVID-19 and its determinants was error-prone. However, there are few studies for the MTUP in crash-frequency modeling. There is still no obvious definition, characteristic, and solution of the MTUP. Previous studies have defined it as consisting of the temporal aggregation, segmentation, and boundary effects or consisting of the duration, temporal resolution, and the point in time aspects [15,16].

In this sense, the purpose of this research is to quantitatively identify the MTUP effects in crash-frequency modeling through a sensitivity analysis using data with various temporal scales. The crash data for 24 highway sections on I-64 of Virginia State from

2011 to 2013 aggregated using four temporal units—a year, a quarter, a month, and a week—were used in the crash-frequency modeling, and the results were compared. The rest of this research is structured with the descriptions of data and the methodology used, followed by a case study and discussions and conclusions for the findings.

2. Data

Crash frequency is the number of crashes of a certain area or a roadway segment during a certain period. In this research, a total of 1827 crash data for 24 highway sections on I-64 of Virginia State from 2011 to 2013 were used, and those included three types of crashes: fatal, injury, and property damage only (PDO). The total number of crashes of each roadway segment was considered in this research instead of division into the three crash types to reduce the number of non-crash observations. Table 1 shows the descriptive statistics of the datasets used in this research.

Table 1. Descriptive Statistics.

Dataset (No. of Obs.)	Variable	Mean	Std. Dev.	Coef. of Variation	Min	Max
Yearly (72)	CRASH_COUNT (/year)	25.4	34.1	1.34	3	185
	SEG_LEN (mile)	3.29	2.09	0.64	0.98	7.38
	ADT (veh/lane/day)	13,060	5231	0.40	1980	21,986
	ADPCP (10 mil/day)	11.1	3.3	0.30	2.5	14.7
Quarterly (288)	CRASH_COUNT (/quarter)	6.3	9.0	1.42	0	64
	SEG_LEN (mile)	3.29	2.08	0.63	0.98	7.38
	ADT (veh/lane/day)	13,085	5356	0.41	1615	23,107
	ADPCP (10 mil/day)	11.5	5.9	0.51	0	31.6
Monthly (864)	CRASH_COUNT (/month)	2.1	3.3	1.54	0	23
	SEG_LEN (mile)	3.29	2.08	0.63	0.98	7.38
	ADT (veh/lane/day)	13,112	5357	0.41	657	23,783
	ADPCP (10 mil/day)	11.9	7.4	0.62	0	37.7
Weekly (3744)	CRASH_COUNT (/week)	0.5	1.0	2.06	0	9
	SEG_LEN (mile)	3.29	2.08	0.63	0.98	7.38
	ADT (veh/lane/day)	13,158	5407	0.41	457	24,764
	ADPCP (10 mil/day)	12.4	14.9	1.20	0	145.7

Four datasets having different temporal scales-yearly, quarterly, monthly, and weekly levels were prepared using the individual crash data. The segment length (SEG_LEN) in miles, average daily traffic (ADT) per lane, and average daily precipitation (ADPCP) in hundredths of inches data were collected as explanatory variables. Since the crash, traffic volume, and precipitation data have time-specific information, they have been aggregated by each temporal scale. On the other hand, the segment length variables in the datasets have the same distributional characteristics, i.e., they do not change over time.

The average daily traffic volume and precipitation data are time-varying variables, i.e., the "average" represents yearly, quarterly, monthly, and weekly averages in the four datasets, respectively. Figure 1 shows the distribution of each variable. While the distributions of the segment length for the four temporal units were identical, for the crash count, traffic volume, and precipitation, the distributions change substantially as the temporal unit changes. In addition, in the yearly averaged dataset, the coefficient variations (CVs) of the crash count and precipitation were relatively smaller than those in the other scaled datasets (see Table 1). This implies that, in the yearly averaged dataset, the variations of those three variables could be less reflected compared to the other scaled datasets. As a result, a bigger information loss could exist in macro temporal unit data than in micro temporal unit data.
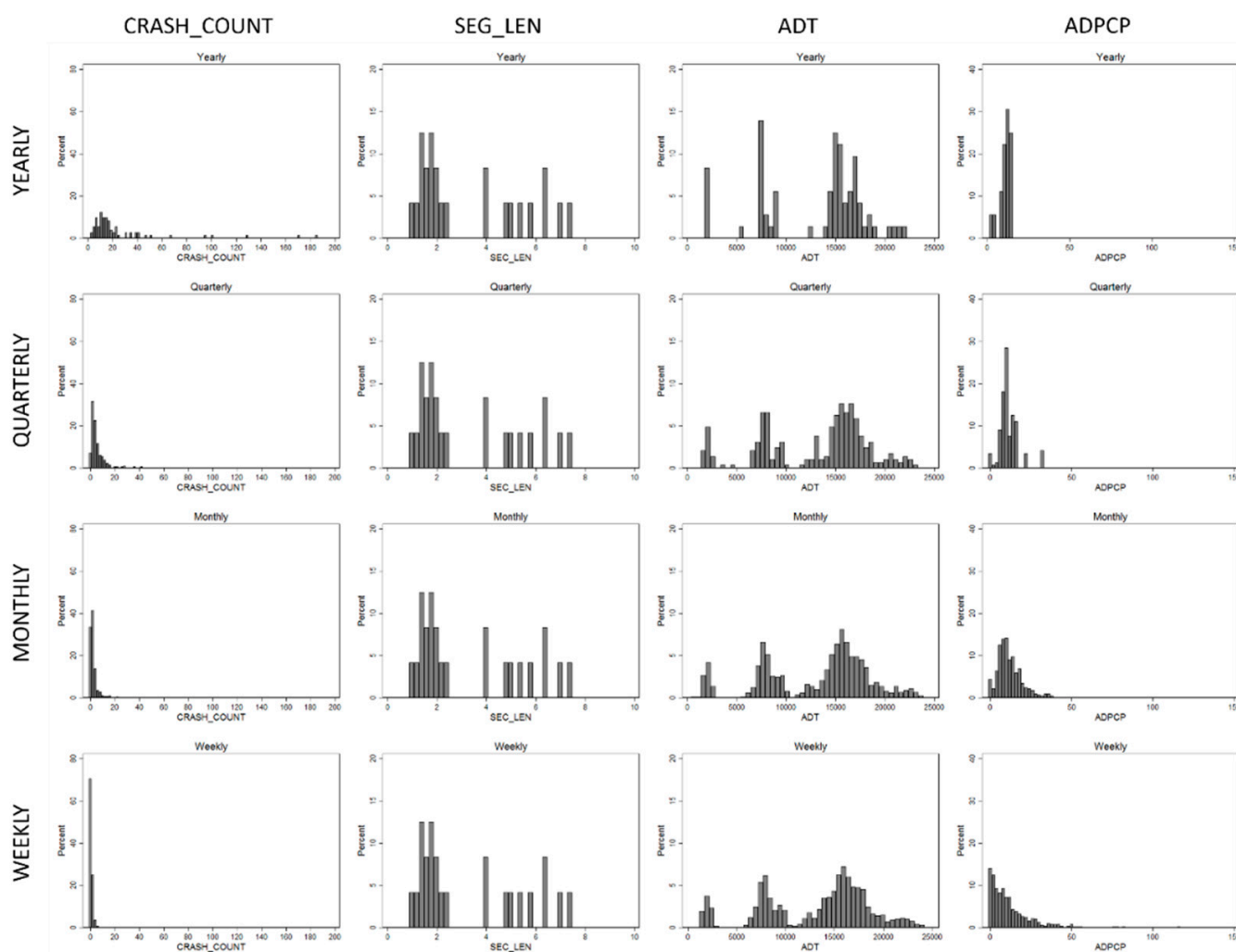
Figure 1. Data Distributions.

3. Analysis Design and Methodology

In order to identify the MTUP effect, a sensitivity analysis framework was adopted using the datasets with four temporal units after the model estimation.

Since crash frequencies on a roadway are discrete and non-negative values, the Poisson regression model and the negative binomial regression model are the most commonly used in crash-frequency modeling. The property of the Poisson distribution is that the mean and variance are equal ($E[y_i] = VAR[y_i]$). The negative binomial model is more appropriate if the data are over dispersed ($E[y_i] < VAR[y_i]$) [17]. When the modifiable temporal units are accounted for, the temporal correlation issue has to be considered because adopting a smaller time interval can make the same observations greater for time-invariant variables, such as a road segment length, the number of lanes, and so on. However, since to identify and compare the effects of different temporal units is also a part of this research, the negative binomial regression model was used first for a crash-frequency model, then the random effect negative binomial models were estimated.

The negative binomial model is derived by the following equation [17]:

$$\lambda_i = EXP(\beta X_i + \varepsilon_i), \tag{1}$$

where λ_i is the mean of the Poisson distribution, X_i is a vector of explanatory variables, β is a vector of estimable parameters, and $EXP(\varepsilon_i)$ is a gamma-distributed error term with

mean 1 and variance α^2. The probability $P(y_i)$ of y_i crashes per a temporal unit of group i is as follows:

$$P(y_i) = \frac{\Gamma((1/\alpha) + y_i)}{\Gamma(1/\alpha)y_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i}\right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i}\right)^{y_i} \tag{2}$$

where $\Gamma(.)$ is a gamma function.

Although crash frequency data used in safety modeling is distributed in time and space, traditional statistical models have mainly been estimated as either cross-sectional or time-series approaches. Unlike cross-sectional models, which suffer from the unobserved time-invariant heterogeneity issue, panel models can take into account variation between groups and within a group, a roadway segment in this research. In addition, as panel data models have higher degrees of freedom and less collinearity among independent variables by increasing observations, more efficient parameter estimates can be derived.

The following equation represents a general form of a panel regression model for a group i and time t:

$$y_{it} = \alpha + X'_{it}\beta + u_i + e_{it} \tag{3}$$

where α is a scalar, X_{it} is the vector of explanatory variables, u_i is the unobserved group-specific effect and e_{it} is the random error. While the fixed effect model regards the group-specific effect, u_i, as a parameter to be estimated, the random effect model considers it as a random variable where $u_i \sim IID(0, \sigma_u^2)$, $e_{it} \sim IID(0, \sigma_e^2)$ and u_i and e_{it} are independent. Therefore, if the groups are randomly sampled from a population and any necessity of considering the group-specific effects exists, the random effect model is appropriate. In this research, as it is reasonable to believe that cross-sectional heterogeneity existed in the crash frequency data and related explanatory variables, the random effect negative binomial model was estimated.

The random effect negative binomial model is as follows [18,19]:

$$\lambda_{it} = EXP(X_{it}\beta + u_i) \tag{4}$$

where u_i is a random effect for the ith group where $EXP(u_i)$ follows a gamma distribution with mean one and variance α. The joint density function is:

$$P(y_{i1}, \cdots, y_{iT}) = \frac{\Gamma(a+b)\Gamma(a + \sum_T \lambda_{it})\Gamma(b + \sum_T y_{it})}{\Gamma(a)\Gamma(b)\Gamma(a+b + \sum_T \lambda_{it} + \sum_T y_{it})} \prod_T \frac{\Gamma(\lambda_{it} + y_{it})}{(\lambda_{it})y_{it}!} \tag{5}$$

where T is the number of time periods, and a and b are the parameters of the underlying beta distribution.

In this research, the MTUP effect was observed in the parameter estimation aspect for crash-frequency models. The parameter estimates of each variable were compared among the four models in terms of the sign, mean, standard error, and significance. In addition, the model fitness was also examined to identify the MTUP effect.

4. Result and Discussion

The four negative binomial regression models were estimated using the yearly, quarterly, monthly, and weekly datasets. Table 2 shows the estimation results.

All the estimates of the dispersion parameter (α in Table 2) for the four models were significantly greater than zero. This implies that the four datasets were over dispersed, and negative binomial models were more appropriate than Poisson models.

For the comparison of goodness-of-fit across models, the mean absolute deviance (MAD) between y_{it} and $\hat{y}_{it}$ And the Akaike Information Criterion (AIC) is reported. Since the MAD could differ by the unit of y_{it}, the MADs for the quarterly, monthly, and weekly models were re-scaled to the yearly metric. Both the pseudo R^2 and MAD indicate that the yearly NB model had a better model fit than other models. Including AIC, all the information criteria based on likelihood function depended on the sample size. Thus, the

AIC values in Table 2 are provided for comparing the goodness-of-fit between the NB and RENB models with the same sample size.

Table 2. Negative Binomial (NB) Model Estimation Results.

Variables	Yearly NB	Quarterly NB	Monthly NB	Weekly NB
SEG_LEN	0.2400 ***	0.2188 ***	0.2227 ***	0.2309 ***
	(0.0390)	(0.0285)	(0.0215)	(0.0164)
ADT	0.1466 ***	0.1282 ***	0.1339 ***	0.1414 ***
	(0.0155)	(0.0103)	(0.0078)	(0.0060)
ADPCP	0.0337	0.0087	0.0011	0.0052 ***
	(0.0231)	(0.0090)	(0.0051)	(0.0017)
Constant	−0.1403	−0.8937 ***	−2.0037 ***	−3.6737 ***
	(0.3488)	(0.2162)	(0.1635)	(0.1291)
α	0.2936 ***	0.5254 ***	0.6580 ***	0.7732 ***
	(0.0543)	(0.0600)	(0.0659)	(0.0816)
Observations, n	72	288	864	3744
Log likelihood	−267.7950	−766.2530	−1538.7473	−3201.5179
Pseudo R^2	0.1225	0.0896	0.0849	0.0828
MAD	12.5843	15.4753	19.0078	30.0578
AIC	545.5900	1542.5060	3087.4946	6413.0358

Standard errors in parentheses. *** $p < 0.01$.

The segment length and average daily traffic in the four NB models were statistically significant at a 99% significance level, having a reasonable sign which was positive. However, the precipitation was significant only for the weekly NB model (See Table 2). It implies that using a small temporal unit can reduce aggregation bias for a variable, especially if it has significantly different distributions when it is aggregated by different temporal levels. Thus, a small temporal unit can derive a more efficient estimate for the variable. From the same perspective, the mean values of the estimates of the segment length were similar since the distributions for the four datasets were identical. Although the average daily traffic was time-variant, unlike the segment length, the magnitude of the estimates was similar among the four models. This resulted from fewer distributional differences between the datasets. In contrast, the estimates of the average daily precipitation presented larger differences in which, as the temporal unit became smaller, the estimate had a smaller value, in addition to showing higher significance. Figure 2 shows the comparison result of the parameter estimates' mean and 95% confidence interval. The decreasing confidence intervals due to decreasing the standard errors of all parameter estimates were observed. The change in the estimate of precipitation was in contrast to one of the others, i.e., the overlapped confidence intervals among the four models were relatively small, which means the means of the estimates were significantly different.
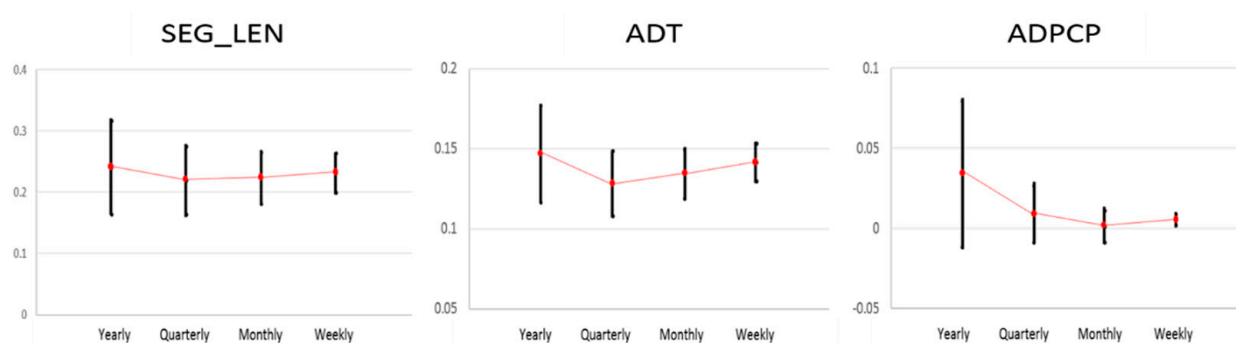

Figure 2. Mean and 95% Confidence Interval of the Estimates in NB models.

As the temporal unit became smaller, the likelihood of temporal correlation increased because the same roadway attributes, e.g., the segment length, were duplicated in multi-

ple observations. Those attributes did not change over time and, consequently, were be correlated to each other due to the remaining same unobserved effects (Lord and Mannering, 2010). Then, the temporal correlation negatively affected the efficiency of parameter estimation. It is known that panel count models, including a panel Poisson and panel negative binomial model, can handle the temporal correlation problem. Among the panel models, the random effect model can account for potential unobserved heterogeneity in the data, as well as the temporal correlation issue. Thus, the random effect negative binomial models using the same four datasets were estimated in order to identify the MTUP effect, excluding the temporal correlation problem. Table 3 represents the estimation results of the random effect models.

Table 3. Random Effect Negative Binomial (RENB) Model Estimation Results.

Variables	Yearly RENB	Quarterly RENB	Monthly RENB	Weekly RENB
SEG_LEN	0.1727 ***	0.0876	0.1062	0.1512 *
	(0.0620)	(0.0669)	(0.0715)	(0.0777)
ADT	0.0001 ***	0.0000 ***	0.0000 ***	0.0000 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ADPCP	−0.0050	0.0063	0.0031	0.0039 **
	(0.0137)	(0.0053)	(0.0037)	(0.0015)
Constant	2.0313 ***	1.3588 ***	1.0750 **	0.3069
	(0.7852)	(0.5207)	(0.4634)	(0.3967)
r	10.8479	7.9595	12.4706	20.2537
	(4.5321)	(2.7254)	(4.4135)	(7.1596)
s	4.6524	3.4821	2.6366	2.1430
	(1.9196)	(1.2471)	(0.8396)	(0.6323)
Observations, n	72	288	864	3744
Number of segments	24	24	24	24
Log likelihood	−238.8837	−680.9758	−1381.7728	−2952.8511
MAD	21.6572	18.7541	22.8054	58.1128
AIC	489.7674	1373.9516	2775.5456	5917.7022

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

For the random effect models, the estimate of the segment length in the yearly RENB model was significant with a significance level of 99%, while it was less significant in the other models (See Table 3). It can be caused by less variability of the variable made by increasing the same observations due to using smaller time intervals. The estimates of the average daily traffic had a high significance in all RE models (See Table 3). Since traffic volume is a well-known major exposure variable in the crash-frequency modeling field, it is reasonable to believe that the variable is normally significant in the models, regardless of the temporal unit and the model type. For the average daily precipitation, the same result with the NB models was observed, i.e., it was significant only for the weekly RENB model (See Table 3).

The four RENB models were superior to the corresponding NB models in terms of the log-likelihood function and AIC (See Table 3). It indicates that the RENB models made improvements in likelihoods by handling unobserved heterogeneity.

Modifying a temporal unit in crash-frequency modeling can influence the significance of the association. It is expected that the standard error of an estimate decreases as the aggregation level of data goes to increasing micro temporal units due to the increasing number of observations. When a smaller temporal unit was used, the magnitude of parameter estimates also decreased due to less variability between the observations (See, Figure 3). These two effects generally cancel out. Hence, the z score remained relatively stable in this research. In general, if the number of observations increases due to differences in temporal units, it would inflate the significance of each estimate, albeit depending on the distributional characteristics of the variables and their aspects of change while adopting

different units. From the statistical perspective, it is desirable since the decreasing standard error of estimates implies more efficient estimates.
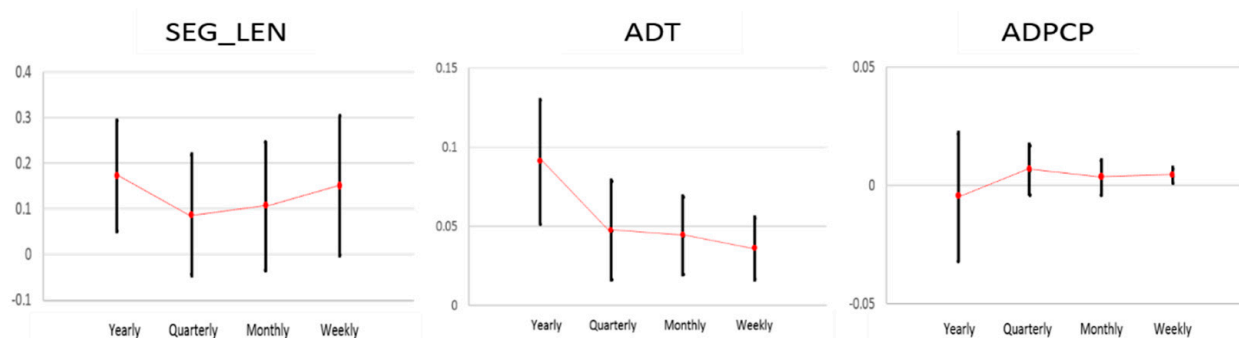

Figure 3. Mean and 95% Confidence Interval of the Estimates in RENB models.

5. Conclusions

This research described the MTUP effect in crash-frequency modeling using four datasets for the same study area, I-64 of Virginia, from 2011 to 2013, and in which the crash data was aggregated by different temporal units: a year, a quarter, a month, and a week. Four negative binomial regression models and four random effect negative binomial models were established with the segment length, average daily traffic, and average daily precipitation as explanatory variables. While the segment length remained the same over the temporal units, the average traffic and precipitation were time-varying variables, which means the data represented the yearly, quarterly, monthly, and weekly averaged values, respectively. The distribution of the precipitation variable substantially changed when different temporal units were used, which means the existence of significant information loss due to the aggregation bias.

As the temporal unit changed, the result of the model estimation also changed in terms of the mean, standard error, and significance of the parameter estimates. For the precipitation variable, the mean and standard error of the parameter estimate decreased, and the statistical significance increased when a smaller temporal unit was used.

For prediction purposes, models based on macro temporal units would be more accurate. However, this does not necessarily mean that they are superior to models having smaller temporal units since the macro models do not provide intuitive findings, i.e., the superiority stems only from statistical aspects. In the assessment of the significance of association, micro models outperform macro models because the standard error of estimates is significantly small, and consequently, more efficient.

Through this research, MTUP was identified in crash-frequency modeling with more effectiveness at a micro-temporal scale. From the policy perspective, using crash-frequency modeling is to identify determinants of crashes and to control them for minimizing possible crashes through proper plans and strategies. Given that there are various factors influencing crashes in a specific period and location, microscopic treatments are required, which have been proven as an effective way to reduce crashes. In this regard, the findings from this research can provide insights to policymakers on how to use the concept of MTUP in crash-frequency modeling and to identify potential determinants for crashes at a certain time and location. This will lead to more tailored countermeasures from relevant policies, which can work better compared to the results from macro-scale models for reducing potential factors on crashes.

Some limitations should be noted. First, the data used in this study were relatively short, allowing only three-year tracking periods for each section of the highway. Since there was limited within-variation in the yearly crash data, fixed effects regression or autoregressive models were unable to be evaluated at different levels of temporal aggregations. Second, the determinants of crash frequency were not fully considered due to limited data

availability. Third, the study site was limited to the I-64 in the state of Virginia. Future study needs to examine whether our findings are generalizable to a broader context or crash data gathered at different settings.

References

1. Miller, H.J. Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geogr. Anal.* **1999**, *31*, 373–399. [CrossRef]
2. Ma, J.M.; Kockelman, K.M.; Damien, P. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accid. Anal. Prev.* **2008**, *40*, 964–975. [CrossRef] [PubMed]
3. Caliendo, C.; Guida, M.; Parisi, A. A Crash-prediction Model for Multilane Roads. *Accid. Anal. Prev.* **2007**, *39*, 657–670. [CrossRef] [PubMed]
4. Ivan, J.N.; Wang, C.Y.; Bernardo, N.R. Explaining Two-lane Highway Crash Rates Using Land Use and Hourly Exposure. *Accid. Anal. Prev.* **2000**, *32*, 787–795. [CrossRef]
5. El-Basyouny, K.; Sayed, T. Collision Prediction Models Using Multivariate Poisson-lognormal Regression. *Accid. Anal. Prev.* **2009**, *41*, 820–828. [CrossRef] [PubMed]
6. Quddus, M.A. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accid. Anal. Prev.* **2008**, *40*, 1486–1497. [CrossRef] [PubMed]
7. Mannering, F.L.; Bhat, C.R. Analytic Methods in Accident Research: Methodological Frontier and Future Directions. *Anal. Methods Accid. Res.* **2014**, *1*, 1–22. [CrossRef]
8. Lord, D.; Mannering, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 291–305. [CrossRef]
9. Lord, D.; Washington, S.P.; Ivan, J.N. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accid. Anal. Prev.* **2005**, *37*, 35–46. [CrossRef] [PubMed]
10. Zhang, M.; Kukadia, N. Metrics of Urban Form and the Modifiable Areal Unit Problem. *Transp. Res. Rec.* **2005**, *1902*, 71–79. [CrossRef]
11. Xu, P.; Huang, H.; Dong, N.; Abdel-Aty, M. Sensitivity Analysis in the Context of Regional Safety Modeling: Identifying and Assessing the Modifiable Areal Unit Problem. *Accid. Anal. Prev.* **2014**, *70*, 110–120. [CrossRef] [PubMed]
12. Jong, R.; Bruin, S. Linear trends in seasonal vegetation time series and the modifiable temporal unit problem. *Biogeosciences* **2012**, *9*, 71–77. [CrossRef]
13. Koch, A.; Carson, D. Spatial, Temporal and Social Scaling in Sparsely Populated Areas—Geospatial Mapping and Simulation Techniques to Investigate Social Diversity. In GI Forum 2012: Geovisualization, Society and Learning. pp. 44–53. Available online: https://www.researchgate.net/publication/261878998_Spatial_Temporal_and_Social_Scaling_in_Sparsely_Populated_Areas_-_geospatial_mapping_and_simulation_techniques_to_investigate_social_diversity (accessed on 31 May 2021).
14. Helbich, M.; Browning, M.H.E.M.; Kwan, M.P. Time to address the spatiotemporal uncertainties in COVID-19 research: Concerns and challenges. *Sci. Total Environ.* **2021**, *764*, 142866. [CrossRef] [PubMed]
15. Cheng, T.; Adepeju, M. Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. *PLoS ONE* **2014**, *9*, e100465.
16. Çöltekin, A.; Sabbata, S.D.; Willi, C.; Vontobe, I.; Pfister, S.; Kuhn, M.; Lacayo, M. *Modifiable Temporal Unit Problem*; ICC 2011 Workshop: Paris, France, 2011.

17. Washington, S.P.; Karlaftis, M.G.; Mannering, F.L. Statistical and Econometric Methods for Transportation Data Analysis. 2003. Available online: https://www.taylorfrancis.com/books/mono/10.1201/9780203497111/statistical-econometric-methods-transportation-data-analysis-simon-washington-matthew-karlaftis-fred-mannering (accessed on 31 May 2021).

18. Shankar, V.N.; Albin, R.B.; Milton, J.C.; Mannering, F.L. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model. *Transp. Res. Rec.* **1998**, *1635*, 44–48. [CrossRef]

19. Hausman, J.; Hall, B.H.; Griliches, Z. Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica* **1984**, *52*, 909–938.