Online Appendix

Understanding Content Contribution Behavior in a Geo-Segmented Mobile Virtual Community: The Context of Waze

Appendix A. Correlation Analysis and Distributions of Boroughs and Weather Conditions

Table A1. Pairwise Correlations between Continuous Variables

Variable	1	2	3	4	5	6	7	8
1. $NumContribution_{it}$	1.000							
2. $PropComment_{it}$	0.171	1.000						
3. $PropThumbUp_{it}$	0.238	-0.071	1.000					
4. $VirtualCrowdedness_{it}$	**0.254**	**0.079**	**0.085**	1.000				
5. $Log(UserSpeed_{it})$	0.075	0.078	0.135	0.105	1.000			
6. $Log(TrafficDen_{i})$	-0.012	0.037	-0.012	0.320	0.057	1.000		
7. $Log(SettledPopDen_{i})$	-0.364	-0.026	-0.096	0.189	-0.129	0.445	1.000	
8. $Log(FloatingPopDen_{i})$	-0.002	-0.008	-0.057	0.445	-0.119	0.571	0.730	1.000
9. $Log(TrafficJamDen_{it})$	0.215	0.050	-0.025	0.462	-0.019	0.297	0.333	0.512
10. $Log(AccidentDen_{it})$	0.055	0.021	0.017	0.111	-0.015	0.069	0.076	0.123
11. $Log(VictimDen_{it})$	0.016	0.010	0.003	0.030	-0.005	0.023	0.029	0.037
12. $DriveCommute_{i}$	0.009	0.016	0.068	-0.387	0.135	-0.445	-0.720	-0.886
13. $Log(CommuteTime_{i})$	-0.176	-0.032	-0.065	-0.361	-0.008	-0.193	-0.040	-0.425
14. $Temperature_{it}$	0.037	0.077	0.032	-0.181	-0.071	0.001	0.002	0.000
15. $Humidity_{it}$	-0.047	-0.036	-0.008	-0.152	-0.056	-0.024	-0.056	-0.065
16. $WindSpeed_{it}$	0.033	0.028	0.049	-0.031	0.072	-0.081	-0.174	-0.234
Variable	9	10	11	12	13	14	15	16
9. $Log(TrafficJamDen_{it})$	1.000							
10. $Log(AccidentDen_{it})$	0.124	1.000						
11. $Log(VictimDen_{it})$	0.039	0.407	1.000					
12. $DriveCommute_{i}$	-0.523	-0.109	-0.037	1.000				
13. $Log(CommuteTime_{i})$	-0.180	-0.076	-0.009	0.218	1.000			
14. $Temperature_{it}$	-0.231	0.028	0.022	0.000	0.004	1.000		
15. $Humidity_{it}$	-0.224	-0.056	-0.019	0.072	0.013	0.236	1.000	
16. $WindSpeed_{it}$	-0.041	-0.017	-0.007	0.259	0.103	-0.263	-0.094	1.000

Table A2. Distributions of Boroughs and Weather Conditions

Borough	Obs.	%	Weather condition	Obs.	%
Bronx	87,024	19.47	Clear	179,857	40.25
Brooklyn	117,600	26.32	Cloudy	229,232	51.30
Manhattan	65,856	14.74	Foggy	3,869	0.87
Queens	134,064	30.00	Rainy	21,473	4.81
Staten Island	42,336	9.47	Snowy	12,449	2.79
Total	446,880	100.00	Total	446,880	100.00

Appendix B. Technical Details for Regression Models

Zero-Inflated Negative Binomial Model. Conditional on the observation not being zero-inflated, it is assumed that the total number of contributions follows a negative binomial distribution. Let y_{it}^* be a latent variable with negative binomial distribution conditional on a set of covariates X_{it}. With the standard negative binomial model setting, the parameter λ_{it} is an exponential function of a vector of covariates: $\lambda_{it} = \exp(X_{it}\beta)$. This paper specifies $X_{it}\beta$ as a linear combination of observed variables. Hence, the probability mass function of y_{it}^*, conditional on λ_{it} and dispersion parameter k can be written as:

$$X_{it}\beta = \alpha_1 + \beta_1 VC_{it} + \beta_2 VC_{it}^2 + \gamma_{11} Z_{1,it} + \gamma_{12} Z_{2,i} + \gamma_{13} Date + \gamma_{14} Hour \qquad \text{(A1)}$$

$$\text{Prob}(y_{it}^* | \lambda_{it}, k) = \frac{\Gamma(y_{it}^* + 1/k)}{\Gamma(y_{it}^* + 1)\Gamma(1/k)} \left(\frac{1}{1 + k\lambda_{it}}\right)^{1/k} \left(\frac{k\lambda_{it}}{1 + k\lambda_{it}}\right)^{y_{it}^*}, \text{for } y_{it}^* = 0,1,2,3,\ldots$$

Notice that total contribution, the dependent variable y_{it}, i.e., $NumContribution_{it}$, equals to zero if the observation is zero-inflated, and y_{it}^* otherwise, as follows:

$$y_{it} = \begin{cases} 0, & \textit{if being excess zero} \\ y_{it}^*, & \textit{otherwise.} \end{cases}$$

The zero-inflation process is modeled using a standard logit model with covariates Z_{it}. Therefore, this paper defines the probability of being zero-inflated as $\pi_{it} = \frac{1}{1+\exp(-Z_{it}\gamma)}$ and specifies $Z_{it}\gamma$ as a linear combination of the observed variables, as follows:

$$Z_{it}\gamma = \alpha_2 + \gamma_{21} Z_{1,it} + \gamma_{22} Z_{2,i} + \gamma_{23} Date + \gamma_{24} Hour \qquad \text{(A2)}$$

Except for VC_{it}, all the covariates are included in the zero-inflation specification, i.e., Equation (A2), in order to model the propensity of having excess zero values for total contribution. Formally, the zero-inflated count data model can be estimated based on a two-stage conditional approach or full information maximum likelihood (Vuong 1989). The full-observation log-likelihood function is derived as follows:

$$LL = \sum_{y_{it}=0} log\left\{\pi_{it} + (1 - \pi_{it})(1 + k\lambda_{it})^{-1/k}\right\} + \sum_{y_{it}>0} \left\{log(1 - \pi_{it}) + log\frac{\Gamma(y_{it}+1/k)}{\Gamma(y_{it}+1)\Gamma(1/k)} + y_{it} log(k\lambda_{it}) - \left(y_{it} + \frac{1}{k}\right) log(1 + k\lambda_{it})\right\} \qquad \text{(A3)}$$

Notice that the first component is the log-likelihood of observing zeros; the second component is the log-likelihood of observing non-zero values.

Heckman Selection Model. The Heckman selection models are specified as follows:

$$y_{it}^{*} = \alpha_1 + \beta_1 VC_{it} + \gamma_{11} Z_{1,it} + \gamma_{12} Z_{2,i} + \gamma_{13} Date + \gamma_{14} Hour + \varepsilon_{it}, \quad \text{(A4)}$$

$$d_{it}^{*} = \alpha_2 + \gamma_{21} Z_{1,it} + \gamma_{22} Z_{2,i} + \gamma_{23} Date + \gamma_{24} Hour + \gamma_{25} Z_{3,i} + \epsilon_{it} \quad \text{(A5)}$$

In the above simultaneous equation model, the first equation is the main stage equation, while the second equation is the selection stage equation. We define that the two error terms jointly follow a bivariate normal distribution. Importantly, the selection process assumes that we could only observe $y_{it} = y_{it}^{*}$ when $d_{it}^{*} > 0$, which means there is at least one total contribution. To meet the exclusion restriction condition, we include additional variables only into the selection equation. These variables, specified by the vector $Z_{3,i}$, include median age of residents (*Age*), proportion of female residents (*Female*), proportion of Caucasian residents (*WhiteRace*), proportion of white-collar residents (*WhiteCollar*), proportion of residents with an associate or higher education degree (*Education*), logarithm of average household income (*LogIncome*), and average yearly spending on cellular phone services per household in USD (*Cellphone*). In Table B1, we show the description and summary statistics of excluded variables.

Table B1. Description and Summary Statistics of Excluded Variables

Variable	Description	Mean	S.D.	Min	Max
Age_i	Median age of residents in NTA *i*	37.02	5.27	17.50	57.25
Female_i	Proportion (×100 %) of female residents in NTA *i*	52.22	3.87	8.41	60.02
WhiteRace_i	Proportion (×100 %) of Caucasian race residents in NTA *i*	55.47	26.51	3.82	98.53
WhiteCollar_i	Proportion (×100 %) of white-collar residents in NTA *i*	60.69	14.83	25.81	95.67
Education_i	Proportion (×100 %) of residents with an associate or higher education degree in NTA *i*	38.62	18.00	4.30	86.99
LogIncome_i	Logarithm of average household income in NTA *i*	11.18	0.42	10.28	12.40
CellPhone_i	Average yearly household spending (in USD) on cellular phone services in NTA *i*	902.55	99.68	625.84	1155.30

Appendix C. Additional Empirical Results

Table C1. Results of Panel Data Models for Total Contribution

Dep. Variable	$NumContribution_{it}$					
Specification	(1) Fixed Effects	(2) Fixed Effects	(3) Pooled OLS	(4) Pooled OLS	(5) Random Effects	(6) Random Effects
VC_{it}	0.9612**	0.5089*	1.4578***	1.0504***	0.9659**	0.5141*
	(0.3551)	(0.2476)	(0.0202)	(0.0211)	(0.3543)	(0.2468)
VC_{it}^2	-0.0109*	-0.0043	-0.0203***	-0.0146***	-0.0110*	-0.0044
	(0.0053)	(0.0039)	(0.0006)	(0.0006)	(0.0053)	(0.0039)
$VC_{it} \times RushHour_t$		1.6977***		1.5996***		1.6969***
		(0.4173)		(0.0522)		(0.4174)
$VC_{it}^2 \times RushHour_t$		-0.0251***		-0.0216***		-0.0250***
		(0.0061)		(0.0018)		(0.0061)
$Log(UserSpeed_{it})$	-1.5986***	-1.4394***	-0.0514	0.0585+	-1.5844***	-1.4252***
	(0.2877)	(0.2493)	(0.0351)	(0.0342)	(0.2881)	(0.2497)
$Log(TrafficDen_i)$			-0.0722	-0.0550	0.4872	0.5174
			(0.0508)	(0.0504)	(0.8319)	(0.8220)
$Log(SettledPopDen_i)$			-23.3313***	-23.4288***	-23.9436***	-24.0799***
			(0.4119)	(0.4092)	(1.4956)	(1.5473)
$Log(FloatingPopDen_i)$			6.2652***	6.5126***	6.3961**	6.7048***
			(0.1706)	(0.1698)	(1.9465)	(1.9489)
$Log(TrafficJamDen_{it})$	5.1735***	4.7052***	6.1570***	5.7966***	5.1812***	4.7140***
	(0.8497)	(0.7302)	(0.0862)	(0.0837)	(0.8498)	(0.7305)
$Log(AccidentDen_{it})$	3.1908*	3.1777*	3.8899***	3.9336***	3.1973*	3.1848*
	(1.5508)	(1.5328)	(0.3325)	(0.3295)	(1.5517)	(1.5336)
$Log(VictimDen_{it})$	-0.6615	-0.6238	-0.6756	-0.6555	-0.6621	-0.6245
	(0.8062)	(0.7887)	(0.5548)	(0.5619)	(0.8059)	(0.7885)
$DriveCommute_i$			-0.2496***	-0.2659***	-0.2893**	-0.3096***
			(0.0078)	(0.0078)	(0.0904)	(0.0910)
$Log(CommuteTime_i)$			7.1740***	6.4722***	5.8524	5.0934
			(0.3696)	(0.3634)	(4.0497)	(3.9643)
$Temperature_{it}$	0.0397**	0.0537***	0.0463***	0.0606***	0.0398**	0.0538***
	(0.0139)	(0.0157)	(0.0094)	(0.0093)	(0.0139)	(0.0157)
$Humidity_{it}$	0.0055	0.0003	0.0086	0.0033	0.0055	0.0003
	(0.0061)	(0.0057)	(0.0061)	(0.0061)	(0.0061)	(0.0057)
$WindSpeed_{it}$	0.0326	0.0421	0.0377***	0.0473***	0.0326	0.0421
	(0.0300)	(0.0318)	(0.0093)	(0.0093)	(0.0300)	(0.0318)
NTA fixed effects	Yes	Yes	No	No	No	No
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Borough dummies	No	No	Yes	Yes	Yes	Yes
Date dummies	Yes	Yes	Yes	Yes	Yes	Yes
Hour-of-day dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs.	446,880	446,880	446,880	446,880	446,880	446,880
Number of NTAs	190	190	190	190	190	190
Adjusted R^2	0.108	0.118	0.391	0.398	0.108	0.118

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and hour-of-day indicators are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C2. Results of Panel Data Negative Binomial Models for Total Contribution

Dep. Variable	$NumContribution_{it}$					
Specification	(1) Pooled NB	(2) Pooled NB	(3) FE NB	(4) FE NB	(5) RE NB	(6) RE NB
VC_{it}	0.0996***	0.0933***	0.0428***	0.0378***	0.0459***	0.0411***
	(0.0008)	(0.0009)	(0.0006)	(0.0007)	(0.0007)	(0.0007)
VC_{it}^2	-0.0017***	-0.0015***	-0.0007***	-0.0005***	-0.0007***	-0.0006***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$VC_{it} \times RushHour_t$		0.0222***		0.0184***		0.0161***
		(0.0015)		(0.0009)		(0.0009)
$VC_{it}^2 \times RushHour_t$		-0.0007***		-0.0005***		-0.0005***
		(0.0001)		(0.0000)		(0.0000)
$Log(UserSpeed_{it})$	0.2435***	0.2449***	-0.0116***	-0.0106***	-0.0257***	-0.0242***
	(0.0028)	(0.0028)	(0.0025)	(0.0025)	(0.0025)	(0.0025)
$Log(TrafficDen_i)$	-0.1149***	-0.1148***			-0.2565***	-0.2561***
	(0.0021)	(0.0021)			(0.0046)	(0.0046)
$Log(SettledPopDen_i)$	-0.2065***	-0.2080***			-0.1606***	-0.1612***
	(0.0016)	(0.0016)			(0.0034)	(0.0034)
$Log(FloatingPopDen_i)$	-0.0584***	-0.0561***			0.0136	0.0151+
	(0.0051)	(0.0051)			(0.0088)	(0.0088)
$Log(TrafficJamDen_{it})$	0.3370***	0.3324***	0.3858***	0.3835***	0.4088***	0.4049***
	(0.0028)	(0.0028)	(0.0026)	(0.0026)	(0.0027)	(0.0028)
$Log(AccidentDen_{it})$	0.2248***	0.2238***	0.1217***	0.1214***	0.1270***	0.1259***
	(0.0131)	(0.0131)	(0.0117)	(0.0117)	(0.0118)	(0.0117)
$Log(VictimDen_{it})$	-0.0325	-0.0317	-0.0121	-0.0105	-0.0006	0.0016
	(0.0306)	(0.0313)	(0.0224)	(0.0223)	(0.0226)	(0.0226)
$DriveCommute_i$	-0.0037***	-0.0039***			-0.0074***	-0.0076***
	(0.0003)	(0.0003)			(0.0004)	(0.0004)
$Log(CommuteTime_i)$	-0.8127***	-0.8175***			-0.4487***	-0.4552***
	(0.0125)	(0.0125)			(0.0214)	(0.0214)
$Temperature_{it}$	0.0030***	0.0031***	0.0010*	0.0010*	0.0020***	0.0020***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
$Humidity_{it}$	0.0006**	0.0005**	0.0008***	0.0008***	0.0001	0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$WindSpeed_{it}$	-0.0006	-0.0005	0.0009*	0.0008*	-0.0007+	-0.0006+
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
NTA fixed effects	No	No	Yes	Yes	No	No
Weather dummies	Yes	Yes	Yes	Yes	Yes	Yes
Borough dummies	Yes	Yes	No	No	No	No
Date dummies	Yes	Yes	Yes	Yes	Yes	Yes
Weekday dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs.	446,880	446,880	446,880	446,880	446,880	446,880
Number of NTAs	190	190	190	190	190	190

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and hour-of-day indicators are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C3. Results of Zero-Inflated Negative Binomial Model for Total Contribution

Dep. Variable	$NumContribution_{it}$	
Specification	(7) ZINB	(8) ZINB
VC_{it}	0.1283***	0.1186***
	(0.0012)	(0.0012)
VC_{it}^2	-0.0021***	-0.0019***
	(0.0000)	(0.0000)
$VC_{it} \times RushHour_t$		0.0380***
		(0.0016)
$VC_{it}^2 \times RushHour_t$		-0.0009***
		(0.0000)
$Log(UserSpeed_{it})$	0.1943***	0.1953***
	(0.0064)	(0.0064)
$Log(TrafficDen_i)$	-0.2752***	-0.2743***
	(0.0065)	(0.0065)
$Log(SettledPopDen_i)$	-0.2680***	-0.2697***
	(0.0048)	(0.0048)
$Log(FloatingPopDen_i)$	-0.1980***	-0.1922***
	(0.0126)	(0.0126)
$Log(TrafficJamDen_{it})$	0.2785***	0.2704***
	(0.0061)	(0.0062)
$Log(AccidentDen_{it})$	0.2251***	0.2282***
	(0.0208)	(0.0208)
$Log(VictimDen_{it})$	0.0350	0.0391
	(0.0419)	(0.0421)
$DriveCommute_i$	-0.0097***	-0.0099***
	(0.0006)	(0.0006)
$Log(CommuteTime_i)$	-0.9516***	-0.9563***
	(0.0258)	(0.0258)
$Temperature_{it}$	0.0049***	0.0050***
	(0.0012)	(0.0012)
$Humidity_{it}$	0.0032***	0.0032***
	(0.0009)	(0.0009)
$WindSpeed_{it}$	-0.0009	-0.0009
	(0.0008)	(0.0008)
NTA fixed effects	No	No
Weather dummies	Yes	Yes
Borough dummies	Yes	Yes
Date dummies	Yes	Yes
Weekday dummies	Yes	Yes
Number of Obs.	446,880	446,880
Number of NTAs	190	190

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and hour-of-day indicators are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C4. Results of Panel Data Models for Feedback and Confirmative Contribution

Dep. Variable	$PropComment_{it}$			$PropThumbUp_{it}$		
Specification	(1) Fixed Effects	(2) Pooled OLS	(3) Random Effects	(4) Fixed Effects	(5) Pooled OLS	(6) Random Effects
VC_{it}	0.1201***	0.2359***	0.1267***	0.0916***	0.4272***	0.1226***
	(0.0117)	(0.0051)	(0.0121)	(0.0139)	(0.0086)	(0.0145)
$Log(UserSpeed_{it})$	-0.4423***	1.5275***	-0.3293***	-0.0286	3.8370***	0.4046***
	(0.0645)	(0.0414)	(0.0650)	(0.0898)	(0.0744)	(0.1053)
$Log(TrafficDen_i)$		0.3935***	0.7466+		0.1572*	1.0098
		(0.0451)	(0.4477)		(0.0762)	(0.8226)
$Log(SettledPopDen_i)$		0.0441+	-0.2732		-1.0747***	-1.8739
		(0.0251)	(0.5829)		(0.0457)	(1.1942)
$Log(FloatingPopDen_i)$		-2.0695***	-2.4979*		-3.7195***	-4.9771**
		(0.0824)	(1.0461)		(0.1396)	(1.6433)
$Log(TrafficJamDen_{it})$	1.6739***	0.7984***	1.6386***	-0.0579	-0.9311***	-0.1034
	(0.1228)	(0.0451)	(0.1237)	(0.1355)	(0.0774)	(0.1401)
$Log(AccidentDen_{it})$	-0.1091	-0.0423	-0.1009	1.4588***	2.5411***	1.5524***
	(0.2223)	(0.2438)	(0.2227)	(0.3769)	(0.4013)	(0.3778)
$Log(VictimDen_{it})$	0.8484+	0.6470	0.8373+	-0.7634	-1.1189	-0.7952
	(0.4564)	(0.4788)	(0.4569)	(0.7684)	(0.7651)	(0.7700)
$DriveCommute_i$		-0.0167***	0.0058		-0.0951***	-0.1124
		(0.0041)	(0.0512)		(0.0069)	(0.0741)
$Log(CommuteTime_i)$		-2.7137***	-3.3956		-10.6142***	-13.9987***
		(0.2125)	(2.5876)		(0.3503)	(3.9485)
$Temperature_{it}$	0.0267**	0.0343***	0.0271**	-0.0210	-0.0068	-0.0196
	(0.0096)	(0.0079)	(0.0095)	(0.0152)	(0.0148)	(0.0152)
$Humidity_{it}$	-0.0120***	-0.0085**	-0.0118***	0.0063	0.0145**	0.0072
	(0.0035)	(0.0030)	(0.0035)	(0.0054)	(0.0055)	(0.0054)
$WindSpeed_{it}$	-0.0232**	-0.0268***	-0.0234**	-0.0151	-0.0203+	-0.0157
	(0.0078)	(0.0068)	(0.0079)	(0.0137)	(0.0120)	(0.0137)
NTA fixed effects	Yes	No	No	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Borough dummies	No	Yes	Yes	No	Yes	Yes
Date dummies	Yes	Yes	Yes	Yes	Yes	Yes
Hour-of-day dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs.	303,473	303,473	303,473	303,473	303,473	303,473
Number of NTAs	190	190	190	190	190	190
Adjusted R^2	0.058	0.068	0.058	0.035	0.066	0.034

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and hour-of-day indicators are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C5. Results of Heckman Selection Model for Feedback and Confirmative Contribution

Dep. Variables	$PropComment_{it}$	$PropThumbUp_{it}$
Specification	(1) Heckman	(2) Heckman
VC_{it}	0.2260***	0.4071***
	(0.0056)	(0.0097)
$Log(UserSpeed_{it})$	1.6433***	3.9210***
	(0.0567)	(0.0976)
$Log(TrafficDen_i)$	0.6040***	0.6814***
	(0.0473)	(0.0814)
$Log(SettledPopDen_i)$	-2.2545***	-5.9722***
	(0.1170)	(0.2013)
$Log(FloatingPopDen_i)$	-1.6028***	-2.6760***
	(0.0876)	(0.1507)
$Log(TrafficJamDen_{it})$	1.2370***	-0.1299
	(0.0680)	(0.1170)
$Log(AccidentDen_{it})$	-0.1301	2.2516***
	(0.2518)	(0.4333)
$Log(VictimDen_{it})$	0.7684	-0.8789
	(0.4705)	(0.8097)
$DriveCommute_i$	-0.0545***	-0.1705***
	(0.0048)	(0.0082)
$Log(CommuteTime_i)$	-1.4369***	-7.4248***
	(0.2226)	(0.3831)
$Temperature_{it}$	0.0339***	-0.0078
	(0.0085)	(0.0146)
$Humidity_{it}$	-0.0093**	0.0135*
	(0.0031)	(0.0054)
$WindSpeed_{it}$	-0.0262***	-0.0187
	(0.0070)	(0.0121)
Inverse Mills ratio	1.3269***	1.6819***
	(0.2937)	(0.5056)
NTA fixed effects	No	No
Weather dummies	Yes	Yes
Borough dummies	Yes	Yes
Date dummies	Yes	Yes
Hour-of-day dummies	Yes	Yes
Number of Obs.	442,176	442,176
Number of NTAs	190	190

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and hour-of-day indicators are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C6. Results of Panel Data Models for Total Contribution (3-Hour Time Window)

Dep. Variable	$NumContribution(perHour)_{it}$					
Specification	(1) Fixed Effects	(2) Fixed Effects	(3) Pooled OLS	(4) Pooled OLS	(5) Random Effects	(6) Random Effects
VC_{it}	1.4724**	0.9219*	1.9945***	1.5133***	1.4846**	0.9354*
	(0.5423)	(0.4327)	(0.0387)	(0.0393)	(0.5376)	(0.4281)
VC_{it}^2	-0.0131	-0.0056	-0.0269***	-0.0213***	-0.0134+	-0.0059
	(0.0080)	(0.0074)	(0.0015)	(0.0013)	(0.0079)	(0.0073)
$VC_{it} \times RushHour_t$		1.6145***		1.4406***		1.6121***
		(0.2804)		(0.0905)		(0.2807)
$VC_{it}^2 \times RushHour_t$		-0.0015		0.0061		-0.0014
		(0.0104)		(0.0054)		(0.0104)
$Log(UserSpeed_{it})$	-2.9991***	-2.7634***	0.1783*	0.3153***	-2.9341***	-2.6990***
	(0.3033)	(0.2691)	(0.0799)	(0.0774)	(0.3041)	(0.2694)
$Log(TrafficDen_i)$			-0.5373***	-0.4900***	0.3716	0.4416
			(0.0767)	(0.0759)	(1.0260)	(1.0179)
$Log(SettledPopDen_i)$			-22.8211***	-22.9420***	-23.4923***	-23.6905***
			(0.6144)	(0.6047)	(1.4295)	(1.5100)
$Log(FloatingPopDen_i)$			5.8548***	6.1636***	5.7148**	6.1542**
			(0.2300)	(0.2272)	(1.9607)	(1.9670)
$Log(TrafficJamDen_{it})$	4.2912***	3.3455***	5.8372***	5.1991***	4.3097***	3.3693***
	(0.7173)	(0.5818)	(0.1641)	(0.1573)	(0.7197)	(0.5836)
$Log(AccidentDen_{it})$	4.5409	4.4968	5.3921***	5.4271***	4.5603	4.5178
	(3.0687)	(2.9877)	(0.6728)	(0.6609)	(3.0749)	(2.9940)
$Log(VictimDen_{it})$	-1.4271	-1.2444	-1.4307	-1.2634	-1.4296	-1.2471
	(1.2465)	(1.1906)	(0.9562)	(0.9813)	(1.2458)	(1.1901)
$DriveCommute_i$			-0.2207***	-0.2461***	-0.2598**	-0.2943**
			(0.0105)	(0.0105)	(0.0908)	(0.0917)
$Log(CommuteTime_i)$			8.6057***	7.6485***	7.4046+	6.3700
			(0.5229)	(0.5030)	(4.3197)	(4.1742)
$Temperature_{it}$	0.0400*	0.0548**	0.0466***	0.0618***	0.0401**	0.0549**
	(0.0156)	(0.0179)	(0.0139)	(0.0136)	(0.0156)	(0.0179)
$Humidity_{it}$	0.0039	-0.0036	0.0070	-0.0011	0.0040	-0.0036
	(0.0060)	(0.0051)	(0.0062)	(0.0061)	(0.0060)	(0.0051)
$WindSpeed_{it}$	0.0536	0.0702	0.0671***	0.0847***	0.0538	0.0705
	(0.0439)	(0.0472)	(0.0161)	(0.0159)	(0.0439)	(0.0472)
NTA fixed effects	Yes	Yes	No	No	No	No
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Borough dummies	No	No	Yes	Yes	Yes	Yes
Date dummies	Yes	Yes	Yes	Yes	Yes	Yes
3-Hour-Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs.	148,580	148,580	148,580	148,580	148,580	148,580
Number of NTAs	190	190	190	190	190	190
Adjusted R^2	0.171	0.197	0.517	0.530	0.171	0.197

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and 3-hour time dummies are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C7. Results of Panel Data Models for Feedback and Confirmative Contribution (3-Hour Time Window)

Dep. Variable	$PropComment_{it}$			$PropThumbUp_{it}$		
Specification	(1) Fixed Effects	(2) Pooled OLS	(3) Random Effects	(4) Fixed Effects	(5) Pooled OLS	(6) Random Effects
VC_{it}	0.2150***	0.4351***	0.2282***	0.1665***	0.8348***	0.2342***
	(0.0269)	(0.0091)	(0.0279)	(0.0282)	(0.0152)	(0.0288)
$Log(UserSpeed_{it})$	-1.1061***	2.8437***	-0.8997***	0.9260***	8.9955***	1.7881***
	(0.1316)	(0.0743)	(0.1345)	(0.1889)	(0.1395)	(0.2067)
$Log(TrafficDen_i)$		0.2974***	0.7548+		-0.0850	1.1668
		(0.0582)	(0.4216)		(0.1018)	(0.8602)
$Log(SettledPopDen_i)$		0.1237***	-0.1199		-0.4768***	-1.3740
		(0.0351)	(0.5686)		(0.0661)	(1.2361)
$Log(FloatingPopDen_i)$		-1.8227***	-2.6875*		-4.1158***	-5.6851**
		(0.1077)	(1.0654)		(0.1887)	(1.8195)
$Log(TrafficJamDen_{it})$	1.4766***	0.1217	1.4132***	-0.4029	-1.8959***	-0.5465+
	(0.1966)	(0.0751)	(0.1984)	(0.2959)	(0.1346)	(0.2946)
$Log(AccidentDen_{it})$	-0.5641	-0.3457	-0.5386	0.4019	2.5086**	0.6103
	(0.5448)	(0.4599)	(0.5440)	(0.8120)	(0.7641)	(0.8084)
$Log(VictimDen_{it})$	0.6814	0.3228	0.6585	1.9423	1.5647	1.8741
	(0.7262)	(0.7793)	(0.7259)	(1.3301)	(1.3627)	(1.3318)
$DriveCommute_i$		-0.0224***	0.0105		-0.1320***	-0.1233
		(0.0053)	(0.0518)		(0.0092)	(0.0814)
$Log(CommuteTime_i)$		-1.6125***	-2.7466		-11.2641***	-15.2395***
		(0.2721)	(2.6603)		(0.4653)	(4.3834)
$Temperature_{it}$	0.0152	0.0196+	0.0154	-0.0175	-0.0067	-0.0163
	(0.0113)	(0.0101)	(0.0113)	(0.0198)	(0.0199)	(0.0198)
$Humidity_{it}$	-0.0147***	-0.0069+	-0.0143***	-0.0048	0.0115	-0.0031
	(0.0040)	(0.0039)	(0.0040)	(0.0063)	(0.0074)	(0.0063)
$WindSpeed_{it}$	-0.0365**	-0.0402***	-0.0367***	0.0104	0.0067	0.0098
	(0.0110)	(0.0102)	(0.0110)	(0.0194)	(0.0189)	(0.0194)
NTA fixed effects	Yes	No	No	Yes	No	No
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Borough dummies	No	Yes	Yes	No	Yes	Yes
Date dummies	Yes	Yes	Yes	Yes	Yes	Yes
3-Hour-Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs.	128,529	128,529	128,529	128,529	128,529	128,529
Number of NTAs	190	190	190	190	190	190
Adjusted R^2	0.086	0.110	0.086	0.067	0.133	0.066

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and 3-hour time dummies are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C8. Results of Panel Data Models for Total Contribution (with Virtual Crowdedness Dummies)

Dep. Variable	$NumContribution_{it}$					
Specification	(1) Fixed Effects	(2) Fixed Effects	(3) Pooled OLS	(4) Pooled OLS	(5) Random Effects	(6) Random Effects
$VC\text{-}10_{it}$	2.2708***	0.5183	4.4324***	2.8209***	2.2901***	0.5415
	(0.4730)	(0.3794)	(0.1336)	(0.1474)	(0.4715)	(0.3790)
$VC\text{-}15_{it}$	8.6919+	4.9818	11.9212***	8.2721***	8.7199+	5.0145
	(4.6923)	(3.5942)	(0.2916)	(0.3193)	(4.6871)	(3.5884)
$VC\text{-}20_{it}$	10.7641*	5.1711+	13.6142***	8.0407***	10.7906*	5.2017*
	(4.7794)	(2.6378)	(0.4704)	(0.4094)	(4.7724)	(2.6296)
$VC\text{-}25_{it}$	11.2904**	5.8302*	14.4670***	9.1005***	11.3199**	5.8647*
	(3.7402)	(2.3332)	(0.5934)	(0.4991)	(3.7330)	(2.3254)
$VC\text{-}30_{it}$	12.1954**	6.8721**	15.8914***	10.8161***	12.2291**	6.9125**
	(3.8688)	(2.5487)	(0.8146)	(0.6889)	(3.8604)	(2.5393)
$VC\text{-}35_{it}$	15.5316**	9.3022**	19.1654***	12.7597***	15.5651**	9.3389**
	(5.2336)	(3.1884)	(1.2353)	(1.0383)	(5.2246)	(3.1786)
$VC\text{-}40_{it}$	17.6502**	9.9140**	21.6929***	13.7783***	17.6869**	9.9546***
	(5.8319)	(3.0197)	(1.6609)	(1.1488)	(5.8227)	(3.0102)
$VC\text{-}45_{it}$	14.1424***	8.1564***	18.8550***	12.8956***	14.1840***	8.2040***
	(3.0430)	(2.2081)	(1.4304)	(1.3375)	(3.0350)	(2.1995)
$VC\text{-}50_{it}$	16.4224***	10.5042***	20.5622***	14.1547***	16.4601***	10.5432***
	(3.3230)	(2.6848)	(1.7871)	(1.8042)	(3.3147)	(2.6743)
$VC\text{-}10_{it} \times RushHour_t$		7.5295***		7.0668***		7.5257***
		(1.2985)		(0.3165)		(1.2985)
$VC\text{-}15_{it} \times RushHour_t$		14.7194**		15.0177***		14.7213***
		(4.4682)		(0.8129)		(4.4694)
$VC\text{-}20_{it} \times RushHour_t$		22.3841**		23.0341***		22.3886**
		(8.5855)		(1.5653)		(8.5875)
$VC\text{-}25_{it} \times RushHour_t$		20.8882***		21.3806***		20.8920***
		(5.4554)		(1.8987)		(5.4574)
$VC\text{-}30_{it} \times RushHour_t$		20.2796***		20.2360***		20.2798***
		(5.1155)		(2.6353)		(5.1173)
$VC\text{-}35_{it} \times RushHour_t$		24.3943**		26.2226***		24.4099**
		(8.0424)		(4.0475)		(8.0433)
$VC\text{-}40_{it} \times RushHour_t$		30.0847*		31.9834***		30.0998*
		(11.7648)		(5.7776)		(11.7666)
$VC\text{-}45_{it} \times RushHour_t$		22.5796***		23.6333***		22.5888***
		(4.7553)		(4.1144)		(4.7563)
$VC\text{-}50_{it} \times RushHour_t$		25.5153***		29.6273***		25.5488***
		(3.6705)		(4.8496)		(3.6706)
$Log(UserSpeed_{it})$	-1.3059***	-1.2394***	0.5883***	0.6145***	-1.2887***	-1.2202***
	(0.1890)	(0.1743)	(0.0330)	(0.0326)	(0.1889)	(0.1742)
$Log(TrafficDen_i)$			0.2121***	0.1981***	0.7309	0.7190
			(0.0510)	(0.0506)	(0.8086)	(0.8026)
$Log(SettledPopDen_i)$			-23.5535***	-23.5895***	-24.0963***	-24.1512***
			(0.4117)	(0.4084)	(1.4533)	(1.4589)
$Log(FloatingPopDen_i)$			6.3119***	6.5166***	6.3589**	6.5943***
			(0.1696)	(0.1692)	(1.9816)	(1.9678)
$Log(TrafficJamDen_{it})$	5.3550***	5.0885***	6.5137***	6.2910***	5.3644***	5.0997***
	(0.8712)	(0.7976)	(0.0867)	(0.0842)	(0.8713)	(0.7978)
$Log(AccidentDen_{it})$	3.2122*	3.2338*	4.0796***	4.1540***	3.2198*	3.2429*
	(1.5697)	(1.5453)	(0.3355)	(0.3331)	(1.5706)	(1.5462)
$Log(VictimDen_{it})$	-0.6527	-0.6460	-0.6909	-0.7156	-0.6534	-0.6470
	(0.8111)	(0.7976)	(0.5587)	(0.5660)	(0.8108)	(0.7973)

$DriveCommute_i$			-0.2829***	-0.2890***	-0.3152***	-0.3229***
			(0.0079)	(0.0078)	(0.0928)	(0.0928)
$Log(CommuteTime_i)$			6.1112***	5.4796***	5.0891	4.4537
			(0.3715)	(0.3648)	(4.0524)	(3.9572)
$Temperature_{it}$	0.0381**	0.0509**	0.0438***	0.0571***	0.0382**	0.0510**
	(0.0145)	(0.0164)	(0.0094)	(0.0093)	(0.0145)	(0.0164)
$Humidity_{it}$	0.0035	-0.0007	0.0065	0.0021	0.0036	-0.0007
	(0.0059)	(0.0057)	(0.0061)	(0.0061)	(0.0059)	(0.0057)
$WindSpeed_{it}$	0.0368	0.0444	0.0436***	0.0514***	0.0369	0.0444
	(0.0312)	(0.0326)	(0.0094)	(0.0093)	(0.0312)	(0.0326)
NTA fixed effects	Yes	Yes	No	No	No	No
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Borough dummies	No	No	Yes	Yes	Yes	Yes
Date dummies	Yes	Yes	Yes	Yes	Yes	Yes
Hour-of-day dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of Obs.	148,580	148,580	148,580	148,580	148,580	148,580
Number of NTAs	190	190	190	190	190	190
Adjusted R^2	0.106	0.113	0.387	0.391	0.106	0.113

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and hour-of-day indicators are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. Time-invariant control variables include $Log(TrafficDen_i)$, $Log(SettledPopDen_i)$, $Log(FloatingPopDen_i)$, $DriveCommute_i$, and $Log(CommuteTime_i)$. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, + $p<0.1$

Appendix D. Details of Matching Estimations

The purpose of matching is to generate comparable groups of observations. We follow existing literature (Haviland et al. 2007, Xu et al. 2017) to conduct a balance diagnosis for treatment effects by comparing the standardized bias before and after matching, as shown in Table D1. Standardized bias is calculated as follows:

$$Std.Bias = \left|M_{X_t} - M_{X_c}\right| / \sqrt{0.5 \times \left(s_{X_t}^2 + s_{X_c}^2\right)}$$

where M_{X_t} and M_{X_c} are the means of a variable X in the treatment group and control group, respectively. To obtain the criterion, the absolute difference of X between the two group is standardized by the pooled standard deviation of the variable X in the two groups, which is calculated as the square root of the average between the variances of X in the treatment group ($s_{X_t}^2$) and the control group ($s_{X_c}^2$). In principle, a smaller standardized bias implies the variable values are more balanced. We calculate standardized bias for each matching variable using the original sample before matching and the matched sample to see whether it has been largely reduced. The comparison provides evidence that the CEM matching improves distributional balance between the treatment group and the control group.

Table D1. Balance Check Results: Coarsened Exact Matching

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3		Group 5 vs. Group 4	
Measure	Std. Bias		Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched	Raw	Matched
$\mathrm{Log(UserSpeed_{it})}$	-0.230	0.016	-0.205	0.005	-0.204	0.017	-0.239	-0.015
$\mathrm{Log(TrafficDen_i)}$	0.913	-0.017	0.371	0.003	0.252	-0.017	0.348	-0.017
$\mathrm{Log(SettledPopDen_i)}$	0.269	-0.013	0.211	-0.011	0.120	-0.010	0.269	-0.032
$\mathrm{Log(FloatingPopDen_i)}$	1.079	0.017	0.387	0.001	0.408	0.024	0.387	0.018
$\mathrm{Log(TrafficJamDen_{it})}$	1.146	0.047	0.486	0.015	0.348	0.024	0.326	0.026
$\mathrm{Log(AccidentDen_{it})}$	0.097	0.006	0.055	0.013	-0.032	0.010	0.047	-0.001
$\mathrm{Log(VictimDen_{it})}$	0.012	0.018	0.013	0.055	-0.056	0.048	0.017	0.051
$\mathrm{DriveCommute_i}$	-1.007	-0.004	-0.422	-0.005	-0.378	-0.006	-0.512	-0.051
$\mathrm{Log(CommuteTime_i)}$	-0.958	-0.026	-0.347	-0.011	-0.418	-0.030	-0.454	-0.048
$\mathrm{Temperature_{it}}$	-0.436	0.030	-0.451	0.018	-0.182	0.002	-0.057	0.002
$\mathrm{Humidity_{it}}$	-0.343	-0.008	-0.215	0.008	-0.157	0.013	-0.203	0.016
$\mathrm{WindSpeed_{it}}$	-0.124	0.003	-0.031	-0.002	-0.035	0.022	-0.119	0.022

Table D2. Balance Check Results: Propensity Score Matching

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3		Group 5 vs. Group 4	
Measure	Std. Bias		Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.203	-0.016	-0.111	-0.020	-0.040	0.029	-0.057	0.012
$Log(TrafficDen_i)$	0.876	-0.044	0.294	-0.014	0.158	-0.018	0.236	-0.106
$Log(SettledPopDen_i)$	0.183	-0.027	0.182	-0.007	0.069	-0.002	0.196	-0.074
$Log(FloatingPopDen_i)$	1.019	0.029	0.276	0.034	0.242	0.030	0.224	0.014
$Log(TrafficJamDen_{it})$	1.101	-0.012	0.345	0.005	0.188	0.033	0.132	0.024
$Log(AccidentDen_{it})$	0.193	0.037	0.102	0.017	0.048	0.001	0.090	-0.002
$Log(VictimDen_{it})$	0.048	-0.003	0.034	0.007	0.015	-0.035	0.012	-0.012
$DriveCommute_i$	-0.950	0.011	-0.304	-0.020	-0.214	-0.001	-0.300	-0.028
$Log(CommuteTime_i)$	-0.927	-0.063	-0.217	-0.051	-0.212	-0.028	-0.221	-0.057
$Temperature_{it}$	-0.458	0.039	-0.475	0.028	-0.181	0.009	-0.041	0.025
$Humidity_{it}$	-0.289	0.001	-0.139	0.005	-0.054	0.019	-0.088	-0.116
$WindSpeed_{it}$	-0.102	-0.022	0.023	-0.024	0.039	-0.009	-0.044	0.026

Table D3. Balance Check Results: Nearest Neighbor Matching

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3		Group 5 vs. Group 4	
Measure	Std. Bias		Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.203	-0.054	-0.111	-0.027	-0.040	0.000	-0.057	0.025
$Log(TrafficDen_i)$	0.876	0.005	0.294	0.011	0.158	0.033	0.236	0.041
$Log(SettledPopDen_i)$	0.183	-0.001	0.182	-0.007	0.069	0.002	0.196	0.029
$Log(FloatingPopDen_i)$	1.019	0.011	0.276	-0.004	0.242	-0.001	0.224	-0.025
$Log(TrafficJamDen_{it})$	1.101	0.046	0.345	0.007	0.188	-0.017	0.132	-0.002
$Log(AccidentDen_{it})$	0.193	0.006	0.102	0.014	0.048	0.036	0.090	0.034
$Log(VictimDen_{it})$	0.048	0.002	0.034	0.007	0.015	0.011	0.012	-0.001
$DriveCommute_i$	-0.950	-0.015	-0.304	-0.006	-0.214	0.000	-0.300	-0.018
$Log(CommuteTime_i)$	-0.927	-0.016	-0.217	-0.003	-0.212	-0.006	-0.221	0.008
$Temperature_{it}$	-0.458	-0.001	-0.475	0.001	-0.181	-0.001	-0.041	-0.026
$Humidity_{it}$	-0.289	-0.015	-0.139	0.007	-0.054	0.013	-0.088	-0.012
$WindSpeed_{it}$	-0.102	0.006	0.023	0.007	0.039	0.042	-0.044	0.036

Table D4. Balance Check Results: Coarsened Exact Matching (Non-Rush Hour)

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3	
Measure	Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.247	0.024	-0.221	0.012	-0.204	0.012
$Log(TrafficDen_i)$	0.973	-0.021	0.410	0.002	0.218	-0.017
$Log(SettledPopDen_i)$	0.302	-0.016	0.256	-0.016	0.149	-0.027
$Log(FloatingPopDen_i)$	1.194	0.035	0.423	0.003	0.434	0.029
$Log(TrafficJamDen_{it})$	1.173	0.054	0.469	0.025	0.309	0.024
$Log(AccidentDen_{it})$	0.094	0.007	0.059	0.009	-0.005	0.024
$Log(VictimDen_{it})$	0.008	0.028	0.013	0.049	-0.071	0.061
$DriveCommute_i$	-1.128	-0.015	-0.475	-0.011	-0.405	-0.005
$Log(CommuteTime_i)$	-1.061	-0.046	-0.387	-0.015	-0.426	-0.031
$Temperature_{it}$	-0.365	0.036	-0.442	0.052	-0.189	0.040
$Humidity_{it}$	-0.372	-0.010	-0.258	-0.003	-0.170	-0.001
$WindSpeed_{it}$	-0.175	0.006	-0.051	-0.002	-0.032	0.018

Table D5. Balance Check Results: Coarsened Exact Matching (Rush Hour)

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3	
Measure	Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.151	0.027	-0.215	0.024	-0.274	0.031
$Log(TrafficDen_i)$	0.805	-0.014	0.381	0.002	0.533	-0.008
$Log(SettledPopDen_i)$	0.285	-0.033	0.162	-0.014	0.169	-0.013
$Log(FloatingPopDen_i)$	0.844	0.040	0.424	-0.012	0.513	0.075
$Log(TrafficJamDen_{it})$	0.976	0.009	0.605	0.001	0.536	-0.071
$Log(AccidentDen_{it})$	-0.007	-0.002	-0.002	0.026	-0.065	0.002
$Log(VictimDen_{it})$	-0.025	0.031	0.008	0.064	-0.017	0.133
$DriveCommute_i$	-0.733	-0.012	-0.433	-0.001	-0.469	-0.019
$Log(CommuteTime_i)$	-0.720	-0.042	-0.355	0.000	-0.521	-0.073
$Temperature_{it}$	-0.706	-0.001	-0.511	-0.080	-0.189	-0.053
$Humidity_{it}$	-0.272	0.001	-0.141	0.006	-0.158	0.031
$WindSpeed_{it}$	-0.010	0.012	-0.067	0.024	-0.099	0.044

Table D6. Balance Check Results: Propensity Score Matching (Non-Rush Hour)

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3	
Measure	Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.221	-0.030	-0.124	-0.008	-0.029	-0.006
$Log(TrafficDen_i)$	0.933	-0.043	0.309	-0.013	0.106	0.055
$Log(SettledPopDen_i)$	0.234	-0.015	0.201	-0.024	0.078	0.053
$Log(FloatingPopDen_i)$	1.139	0.039	0.295	0.007	0.255	0.009
$Log(TrafficJamDen_{it})$	1.141	0.007	0.339	0.003	0.177	0.022
$Log(AccidentDen_{it})$	0.208	0.023	0.115	-0.006	0.073	0.015
$Log(VictimDen_{it})$	0.053	0.001	0.031	0.006	0.022	-0.010
$DriveCommute_i$	-1.081	0.009	-0.328	0.004	-0.214	-0.022
$Log(CommuteTime_i)$	-1.029	-0.068	-0.254	-0.022	-0.226	-0.019
$Temperature_{it}$	-0.387	0.029	-0.469	0.039	-0.193	-0.010
$Humidity_{it}$	-0.324	-0.002	-0.183	-0.002	-0.048	-0.005
$WindSpeed_{it}$	-0.153	-0.024	0.023	-0.014	0.061	-0.007

Table D7. Balance Check Results: Propensity Score Matching (Rush Hour)

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3	
Measure	Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.096	0.011	-0.066	0.007	-0.081	-0.062
$Log(TrafficDen_i)$	0.712	-0.078	0.263	-0.002	0.314	0.002
$Log(SettledPopDen_i)$	0.034	-0.159	0.144	-0.021	0.056	0.002
$Log(FloatingPopDen_i)$	0.683	-0.067	0.253	0.009	0.213	0.079
$Log(TrafficJamDen_{it})$	0.888	-0.006	0.375	-0.049	0.258	0.024
$Log(AccidentDen_{it})$	0.132	-0.028	0.062	-0.039	-0.050	0.041
$Log(VictimDen_{it})$	0.028	-0.024	0.041	0.004	-0.002	0.020
$DriveCommute_i$	-0.609	0.063	-0.289	0.004	-0.232	-0.060
$Log(CommuteTime_i)$	-0.621	-0.013	-0.128	-0.023	-0.176	-0.094
$Temperature_{it}$	-0.757	0.065	-0.506	0.009	-0.144	0.067
$Humidity_{it}$	-0.169	-0.002	-0.016	-0.037	-0.068	0.075
$WindSpeed_{it}$	0.043	0.005	0.017	-0.022	-0.013	-0.051

Table D8. Balance Check Results: Nearest Neighbor Matching (Non-Rush Hour)

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3	
Measure	Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.221	-0.043	-0.124	-0.029	-0.029	0.021
$Log(TrafficDen_i)$	0.933	0.002	0.309	0.014	0.106	0.039
$Log(SettledPopDen_i)$	0.234	-0.003	0.201	-0.003	0.078	0.001
$Log(FloatingPopDen_i)$	1.139	0.014	0.295	0.003	0.255	0.002
$Log(TrafficJamDen_{it})$	1.141	0.051	0.339	0.021	0.177	-0.016
$Log(AccidentDen_{it})$	0.208	0.006	0.115	0.015	0.073	0.043
$Log(VictimDen_{it})$	0.053	0.001	0.031	0.002	0.022	0.007
$DriveCommute_i$	-1.081	-0.020	-0.328	-0.012	-0.214	-0.002
$Log(CommuteTime_i)$	-1.029	-0.021	-0.254	-0.008	-0.226	-0.008
$Temperature_{it}$	-0.387	-0.003	-0.469	0.000	-0.193	-0.004
$Humidity_{it}$	-0.324	-0.015	-0.183	0.002	-0.048	0.016
$WindSpeed_{it}$	-0.153	0.012	0.023	0.015	0.061	0.035

Table D9. Balance Check Results: Nearest Neighbor Matching (Rush Hour)

Comparison	Group 2 vs. Group 1		Group 3 vs. Group 2		Group 4 vs. Group 3	
Measure	Std. Bias		Std. Bias		Std. Bias	
Sample	Raw	Matched	Raw	Matched	Raw	Matched
$Log(UserSpeed_{it})$	-0.096	-0.090	-0.066	-0.047	-0.081	-0.066
$Log(TrafficDen_i)$	0.712	0.033	0.263	0.019	0.314	0.090
$Log(SettledPopDen_i)$	0.034	-0.002	0.144	-0.015	0.056	0.004
$Log(FloatingPopDen_i)$	0.683	0.017	0.253	-0.021	0.213	0.026
$Log(TrafficJamDen_{it})$	0.888	0.016	0.375	-0.006	0.258	-0.032
$Log(AccidentDen_{it})$	0.132	0.013	0.062	0.033	-0.050	0.013
$Log(VictimDen_{it})$	0.028	0.004	0.041	0.028	-0.002	0.022
$DriveCommute_i$	-0.609	-0.022	-0.289	-0.007	-0.232	-0.009
$Log(CommuteTime_i)$	-0.621	-0.016	-0.128	0.007	-0.176	-0.033
$Temperature_{it}$	-0.757	-0.003	-0.506	0.000	-0.144	-0.007
$Humidity_{it}$	-0.169	-0.026	-0.016	0.009	-0.068	0.005
$WindSpeed_{it}$	0.043	0.002	0.017	-0.006	-0.013	0.030

Appendix E. Results of Alternative Matching Estimations

In addition to CEM, we also apply two alternative matching estimators, namely *propensity score matching* (PSM) and *nearest-neighbor matching* (NNM). For propensity score matching, we first measure the estimated treatment probability, i.e., propensity score, using a first-stage binary regression model with exactly the same set of covariates used by CEM matching. Then, we utilize the propensity score to derive the matched sample for estimation. For nearest-neighbor matching, we choose Mahalanobis distance to match the observations using the same set of covariates.

Figure E1. Results of Matching Estimations

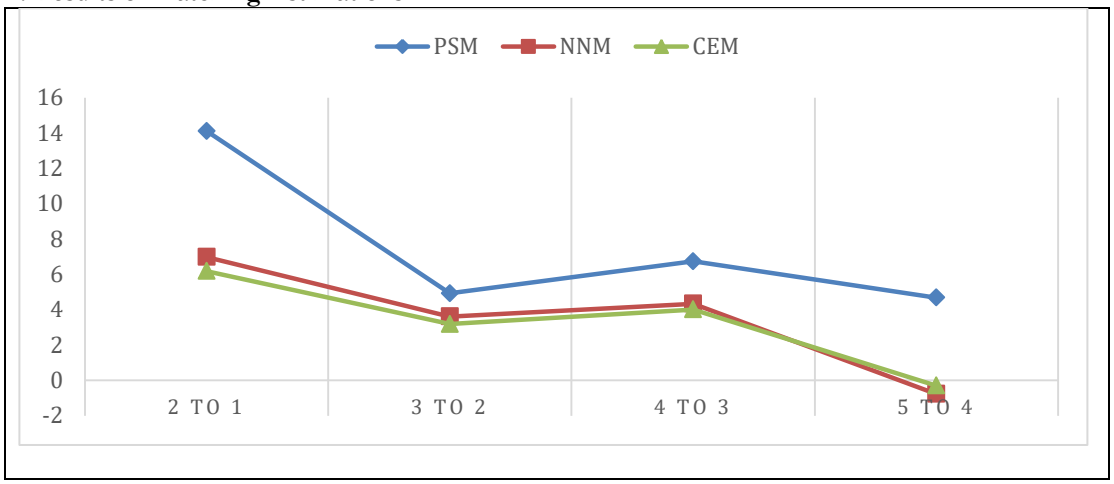


Table E1. Results of Matching Estimations

Matching Method	Treatment groups	Obs.	ATET	standard error	z-value	p-value
PSM	Group 2 vs. Group 1	63,686	14.1084	0.5805	24.31	0.000
	Group 3 vs. Group 2	13,560	4.9243	1.1043	4.46	0.000
	Group 4 vs. Group 3	4,418	6.7379	2.0078	3.36	0.001
	Group 5 vs. Group 4	3,194	4.6832	1.6881	2.77	0.006
NNM	Group 2 vs. Group 1	63,686	6.9829	0.3709	18.82	0.000
	Group 3 vs. Group 2	13,560	3.6143	0.6099	5.93	0.000
	Group 4 vs. Group 3	4,418	4.3373	1.1171	3.88	0.000
	Group 5 vs. Group 4	3,194	-0.7696	1.2580	-0.61	0.541
CEM	Group 2 vs. Group 1	55,290	6.1824	0.3775	16.38	0.000
	Group 3 vs. Group 2	11,192	3.1928	0.6815	4.69	0.000
	Group 4 vs. Group 3	3,340	4.0156	1.2818	3.13	0.002
	Group 5 vs. Group 4	2,292	-0.2976	1.5732	-0.19	0.850

Note. Dependent variable: $y_{it} = NumContribution_{it}$; treatments: the five groups evenly divided along by the distribution of VC_{it} *where* $0 < VC_{it} < 50$. The PSM includes all the control variables as well as the dummy variables: weather dummies, borough indicators, date dummies and hour-of-day dummies. NNM includes all the control variables and dummies but coded date (include an additional dummy for summer/winter) and hour as continuous variables. For CEM, we first perform exact matching based on the quartiles of the continuous controls and categorical controls (including weather dummies), and then choose the closest observation to match based on borough dummies, and continuously-coded date and hour-of-day.

Table E2. Results of Matching Estimations (Non-Rush Hour)

Matching Method	Treatment groups	Obs.	ATET	standard error	z-value	p-value
PSM	Group 2 vs. Group 1	48,706	12.2607	0.5833	20.93	0.000
	Group 3 vs. Group 2	10,280	5.3887	0.8383	6.43	0.000
	Group 4 vs. Group 3	3,382	2.3637	1.6269	1.45	0.146
NNM	Group 2 vs. Group 1	48,706	6.6906	0.3997	16.74	0.000
	Group 3 vs. Group 2	10,280	3.2019	0.5374	5.96	0.000
	Group 4 vs. Group 3	3,382	3.0532	0.9023	3.38	0.000
CEM	Group 2 vs. Group 1	42,620	5.8544	0.4104	14.27	0.000
	Group 3 vs. Group 2	8,522	3.5855	0.6001	5.98	0.000
	Group 4 vs. Group 3	2,570	3.5198	1.1257	3.13	0.002

Note. Dependent variable: $y_{it} = NumContribution_{it}$; treatments: the five groups evenly divided along by the distribution of VC_{it} *where* $0 < VC_{it} < 50$. The PSM includes all the control variables as well as the dummy variables: weather condition dummies, borough indicators, date dummies and hour-of-day dummies. NNM includes all the control variables and dummies but coded date (include an additional dummy for summer/winter) and hour as continuous variables. For CEM, we first perform exact matching based on the quartiles of the continuous controls and categorical controls (including weather dummies), and then choose the closest observation to match based on borough dummies, and continuously coded date and hour-of-day.

Table E3. Results of Matching Estimations (Rush Hour)

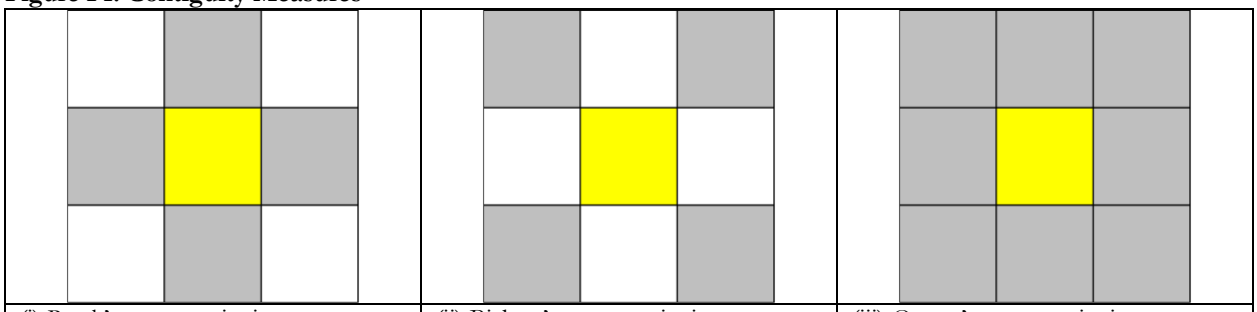
Matching Method	Treatment groups	Obs.	ATET	standard error	z-value	p-value
PSM	Group 2 vs. Group 1	14,980	19.6872	1.7550	11.22	0.000
	Group 3 vs. Group 2	3,280	5.2037	3.6050	1.44	0.149
	Group 4 vs. Group 3	1,036	3.7201	8.7422	0.43	0.670
NNM	Group 2 vs. Group 1	14,980	8.2275	0.7411	11.10	0.000
	Group 3 vs. Group 2	3,280	5.3884	1.4831	3.63	0.000
	Group 4 vs. Group 3	1,036	5.5579	2.9064	1.91	0.056
CEM	Group 2 vs. Group 1	10,382	7.8582	0.8989	8.74	0.000
	Group 3 vs. Group 2	2,094	6.9962	1.8080	3.87	0.000
	Group 4 vs. Group 3	576	5.6806	3.6721	1.55	0.122

Note. Dependent variable: $y_{it} = NumContribution_{it}$; treatments: the five groups evenly divided along by the distribution of VC_{it} *where* $0 < VC_{it} < 50$. The PSM includes all the control variables as well as the dummy variables: weather condition dummies, borough indicators, date dummies and hour-of-day dummies. NNM includes all the control variables and dummies but coded date (include an additional dummy for summer/winter) and hour as continuous variables. For CEM, we first perform exact matching based on the quartiles of the continuous controls and categorical controls (including weather dummies), and then choose the closest observation to match based on borough dummies, and continuously coded date and hour-of-day.

Appendix F. Technical Details for Spatial Models

In spatial analysis, we first need to model the adjacency relationships between observations. To begin with, we need to define a spatial weighting matrix to describe such adjacency relationships depending on the continuity measure. Contiguity refers to how to set boundaries among spatial neighbors. The most commonly used contiguity measures are *rook*'s, *bishop*'s, and *queen*'s case contiguity (Lloyd 2010).[1] As depicted in Figure F1, rook's case contiguity sets the spatial neighbors, which share their common boundaries along *borders* only, as adjacent units whereas bishop's case contiguity recognizes the spatial units, which are adjacent at *points* only, as neighbors. Queen's case contiguity sets the spatial neighbors, which share either of common *borders* or common *points*, as neighboring units.

Figure F1. Contiguity Measures



(i) Rook's case contiguity | (ii) Bishop's case contiguity | (iii) Queen's case contiguity

Note. In each case, the yellow-colored cell is a focal zone whereas the gray-colored cells are spatial neighbors of the focal zone.

We choose queen's case contiguity to generate spatial weight matrix due to the transportation context in which mobile app users are driving along roads in the networks of urban areas. Specifically, the *n*-by-*n* spatial queen's case weight matrix W is expressed as follows:

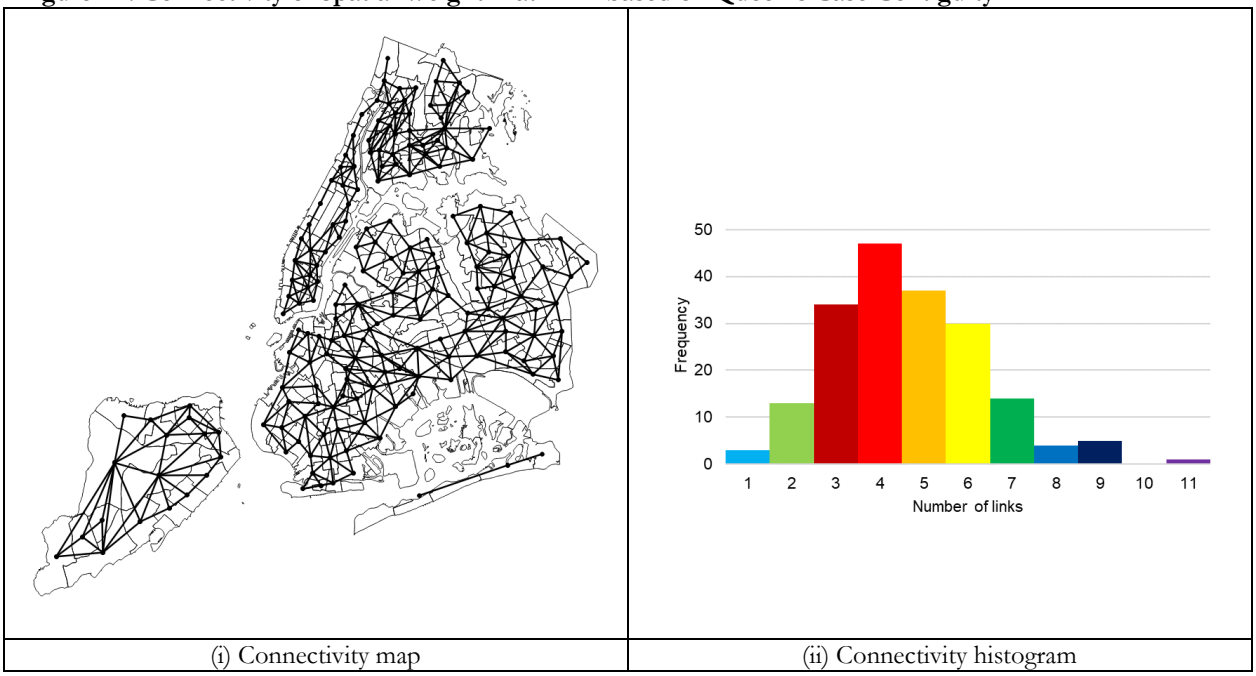
$$W = \begin{pmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nn} \end{pmatrix}$$

where $\omega_{ij} = 1$ if Zone *i* and Zone *j* share at least a boundary or a point; $\omega_{ij} = 0$ if they do not share any boundary or point; $\omega_{ij} = 0$ if $i = j$. In other words, element ω_{ij} of the spatial weight matrix W equals to one if NTA *i* and NTA *j* are adjacent to each other, or equals to zero if NTA *i* and NTA *j* are not adjacent to each other (LeSage and Pace 2009). The diagonal elements ω_{ii} are set to zero, because in principle we do not allow

[1] The names of contiguity measures originate from chess.

spatial effects from themselves. On the left panel of Figure F2, we visualize the neighborhood relationships between 188 NTAs, as each pair of neighboring NTAs are linked with a line. On the right panel, we show the distribution of number of neighbors (links) to the NTAs. The number of neighbors varies from one to eleven, while the average number of neighbors is 4.628.

Figure F2. Connectivity of Spatial Weight Matrix $\boldsymbol{W}$ based on Queen's Case Contiguity



(i) Connectivity map | (ii) Connectivity histogram

After obtaining the row-normalized weighting matrix W, we could start to build spatial models. Spatial models differ from the traditional regression models, in the sense that the error terms are spatially correlated, thus violating the independence (of error) assumption of the simple linear regression model. Based on the previous literature on spatial econometrics (Anselin and Rey 2014; Elhorst 2014), we derive the maximum likelihood (ML) estimation for the spatial models. Using matrix notation, we specify the fixed-effects panel data spatial autoregressive model (SAR) as follows.

$$\boldsymbol{y} = \boldsymbol{\rho}(\boldsymbol{I_T} \otimes \boldsymbol{W})\boldsymbol{y} + \boldsymbol{X\beta} + (\boldsymbol{\tau_T} \otimes \boldsymbol{I_N})\boldsymbol{\mu} + \boldsymbol{\epsilon} \qquad \text{(A6)}^2$$

$$\boldsymbol{\epsilon} \sim \boldsymbol{MVN}(\boldsymbol{0}, \boldsymbol{\sigma^2 I_{NT}}) \qquad \text{(A7)}$$

The spatial weight matrix $\boldsymbol{W}$ is an $N \times N$[3] matrix of spatial weights of NTAs over NYC based on the

[2] $\boldsymbol{I_T}$ is an identity matrix with size T, and symbol $\otimes$ stands for Kronecker product.
[3] N is total number of areas, which equals to 188; T is total number of periods.

queen's case contiguity measure. To deal with panel data, we stack the observations of each cross-section, for $t = 1,2, \dots, T$, to obtain the *NT×1* vector for the dependent variable $\boldsymbol{y}$ as well as for the spatially lagged term $(\boldsymbol{I_T} \otimes \boldsymbol{W})\boldsymbol{y}$. Importantly, $(\boldsymbol{I_T} \otimes \boldsymbol{W})\boldsymbol{y}$ captures a first-order spatial autoregressive process on the focal NTA. $\boldsymbol{X}$ is the $NT \times k$ matrix of exogenous explanatory variables, where k is the number of such variables. Finally, $\boldsymbol{\mu}$ denotes an *N×1* vector of NTA specific fixed effects, so that $(\boldsymbol{\tau_T} \otimes \boldsymbol{I_N})\boldsymbol{\mu}$ is an *NT×1* stacked NTA fixed effects[4]. From equation (A6), we can derive the reduced form regression:

$$\boldsymbol{y} = \left(\boldsymbol{I_{NT}} - \boldsymbol{\rho}(\boldsymbol{I_T} \otimes \boldsymbol{W})\right)^{-1}(\boldsymbol{X\beta} + (\boldsymbol{\tau_T} \otimes \boldsymbol{I_N})\boldsymbol{\mu} + \boldsymbol{\epsilon}) \qquad \text{(A8)}$$

Notice that equation (A8) indicates that the error terms are spatially correlated. To obtain the ML estimate of fixed effects SAR model, we first use the "within transformation" approach to get rid of the fixed effects (as nuisance parameters), by demeaning the y and x variables:

$$y_{it}^* = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} \quad and \quad x_{it}^* = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it} \qquad \text{(A9)}$$

To derive the joint density function of the random variable $\boldsymbol{y}^*$ given the known density function of the error term $\boldsymbol{\epsilon}$, we could use the Jacobian transformation trick. Based on equations (A6) and (A7), we know that the joint density function of random variable $\boldsymbol{\epsilon}$ is: $\boldsymbol{f(\epsilon)} = \frac{1}{\sqrt{(2\pi\sigma^2)^{NT}}} exp\left(\frac{-1}{2\sigma^2}\boldsymbol{\epsilon'\epsilon}\right)$. Notice that the Jacobian transformation implies that given the known density function of *x* as *f(x)*, the density function of *y* is $g(y) = f(x(y))\left|\frac{\partial x}{\partial y}\right|$, where $\left|\frac{\partial x}{\partial y}\right|$ is the Jacobian, which is the determinant of a matrix of partial derivatives. Notice that $\boldsymbol{\epsilon}$ can be written as a function of y: $\boldsymbol{\epsilon} = \boldsymbol{g(y^*)} = \boldsymbol{y^* - \rho(I_T \otimes W)y^* - X^*\beta}$. Therefore, we have a likelihood function:

$$\boldsymbol{f(y^*)} = \left|\frac{\boldsymbol{\partial(y^* - \rho(I_T \otimes W)y^* - X^*\beta)}}{\boldsymbol{\partial y^*}}\right| \boldsymbol{f\left(g(y^*)\right)}$$

$$= |\boldsymbol{I_{NT} - \rho(I_T \otimes W)}| \frac{1}{\sqrt{(2\pi\sigma^2)^{NT}}} exp\left(\frac{-1}{2\sigma^2}(\boldsymbol{y^* - \rho(I_T \otimes W)y^* - X^*\beta})'(\boldsymbol{y^* - \rho(I_T \otimes W)y^* - X^*\beta})\right) \text{(A10)}$$

Given the density function of $\boldsymbol{y}^*$, we can write the corresponding log-likelihood function:

$$log(L) =$$

[4] $\boldsymbol{\tau_T}$ is a T-element column vector of 1.

$$Tlog|\boldsymbol{I_N} - \rho\boldsymbol{W}| - \left(\frac{NT}{2}\right)(log2\pi\sigma^2) - \left(\frac{1}{2\sigma^2}\right)(\boldsymbol{y}^* - \rho(\boldsymbol{I_T}\otimes\boldsymbol{W})\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta})'(\boldsymbol{y}^* - \boldsymbol{\rho}(\boldsymbol{I_T}\otimes\boldsymbol{W})\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta})$$

(A11) [5]

The parameter space is the vector $\boldsymbol{\Theta} = \{\rho, \boldsymbol{\beta}, \sigma^2\}$, with number of parameters equaling to *k+2*. It is hard to conduct a maximum likelihood procedure for all the parameters at the same time. The spatial econometrics literature applies a concentrated log-likelihood function to reduce the computational burden. The ML estimates of $\boldsymbol{\beta}$ and σ^2 can be analytically solved, like the simple linear regression, and therefore they can be written as a function of unknown parameter ρ. To derive the ML estimate of $\boldsymbol{\beta}$, we define:

$$\widehat{\boldsymbol{\beta_0}} = \left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{y}^* \quad \text{(A12)}$$

$$\widehat{\boldsymbol{\beta_1}} = \left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*\prime}(\boldsymbol{I_T}\otimes\boldsymbol{W})\boldsymbol{y}^* \quad \text{(A13)}$$

$$\widehat{\boldsymbol{\beta_{ML}}} = \widehat{\boldsymbol{\beta_0}} - \hat{\rho}\widehat{\boldsymbol{\beta_1}} \quad \text{(A14)}$$

Since $\hat{\rho}$ is an unknown parameter to be estimated, equation (A14) expresses the ML estimates of $\boldsymbol{\beta}$ as a function of $\hat{\rho}$. Practically, $\widehat{\boldsymbol{\beta_0}}$ is the OLS estimate of regressing $\boldsymbol{y}^*$ on $\boldsymbol{X}^*$, while $\widehat{\boldsymbol{\beta_1}}$ is the OLS estimate of regressing $(\boldsymbol{I_T}\otimes\boldsymbol{W})\boldsymbol{y}^*$ on $\boldsymbol{X}^*$. Similarly, we derive the ML estimate of σ^2 using the residuals, as a function of $\hat{\rho}$ as follows.

$$\boldsymbol{e_0} = \boldsymbol{y}^* - \boldsymbol{X}^*\widehat{\boldsymbol{\beta_0}} \quad \text{(A15)}$$

$$\boldsymbol{e_1} = (\boldsymbol{I_T}\otimes\boldsymbol{W})\boldsymbol{y}^* - \boldsymbol{X}^*\widehat{\boldsymbol{\beta_1}} \quad \text{(A16)}$$

$$\widehat{\sigma^2_{ML}} = \frac{1}{NT}(\boldsymbol{e_0} - \hat{\rho}\boldsymbol{e_1})'(\boldsymbol{e_0} - \hat{\rho}\boldsymbol{e_1}) \quad \text{(A17)}$$

Substitute (A12) and (A13) into (A14), and then substitute (A14) and (A17) into (A11), we obtain the concentrated log-likelihood as follows.

$$log(L_C) = Tlog|\boldsymbol{I_N} - \rho\boldsymbol{W}| - \left(\frac{NT}{2}\right)log\left(\frac{1}{NT}(\boldsymbol{e_0} - \hat{\rho}\boldsymbol{e_1})'(\boldsymbol{e_0} - \hat{\rho}\boldsymbol{e_1})\right) + constant \quad \text{(A18)}$$

The ML estimate of ρ can be obtained by numerically solving the single-variable maximization problem based on the concentrated log-likelihood. Finally, we substitute $\widehat{\rho_{ML}}$ into (A14) and (A17) to get the ML estimates of $\boldsymbol{\beta}$ and σ^2. The asymptotic variance matrix of $\boldsymbol{\Theta} = \{\rho, \boldsymbol{\beta}, \sigma^2\}$ can be calculated by using the

[5] Notice that $log|\boldsymbol{I_{NT}} - \boldsymbol{\rho}(\boldsymbol{I_T}\otimes\boldsymbol{W})| = Tlog|\boldsymbol{I_N} - \rho\boldsymbol{W}|$, based on the block diagonal structure of the matrix.

inverse of the information matrix, as follows (Anselin and Rey 2014; Elhorst 2014).

$$\begin{bmatrix} T\cdot tr\left(\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{W}}+\widetilde{\boldsymbol{W}}'\widetilde{\boldsymbol{W}}\right)+\boldsymbol{\beta}'\,\boldsymbol{X}^{*\prime}\left(\boldsymbol{I_T}\otimes\widetilde{\boldsymbol{W}}'\widetilde{\boldsymbol{W}}\right)\boldsymbol{X}^{*}\boldsymbol{\beta}/\sigma^2 & \boldsymbol{X}^{*\prime}\left(\boldsymbol{I_T}\otimes\widetilde{\boldsymbol{W}}\right)\boldsymbol{X}^{*}\boldsymbol{\beta}/\sigma^2 & T\cdot tr\left(\widetilde{\boldsymbol{W}}\right)/\sigma^2 \\ \boldsymbol{X}^{*\prime}\left(\boldsymbol{I_T}\otimes\widetilde{\boldsymbol{W}}\right)\boldsymbol{X}^{*}\boldsymbol{\beta}/\sigma^2 & \boldsymbol{X}^{*\prime}\boldsymbol{X}^{*}/\sigma^2 & 0 \\ T\cdot tr\left(\widetilde{\boldsymbol{W}}\right)/\sigma^2 & 0 & NT/2\sigma^4 \end{bmatrix}^{-1}$$

$$where\ \widetilde{\boldsymbol{W}}=\boldsymbol{W}(\boldsymbol{I_N}-\rho\boldsymbol{W})^{-\mathbf{1}}$$

(A19)[6]

Finally, to derive the direct and indirect marginal effects of the *k-th* explanatory variable on the dependent variable, we could take partial derivatives of the expected value of $\boldsymbol{y}$ with respect to the *k*-th explanatory variable of $\boldsymbol{X}$ from each observation. Given the reduced-form regression equation $\boldsymbol{y}=\left(\boldsymbol{I_{NT}}-\boldsymbol{\rho}(\boldsymbol{I_T}\otimes\boldsymbol{W})\right)^{-\mathbf{1}}(\boldsymbol{X\beta}+(\boldsymbol{\tau_T}\otimes\boldsymbol{I_N})\boldsymbol{\mu}+\boldsymbol{\epsilon})$, we could derive the partial derivative matrix:

$$S_k(\boldsymbol{I_T}\otimes\boldsymbol{W})=\left(\boldsymbol{I_{NT}}-\boldsymbol{\rho}(\boldsymbol{I_T}\otimes\boldsymbol{W})\right)^{-\mathbf{1}}\boldsymbol{I_{NT}}\boldsymbol{\beta}_k \tag{A20}$$

Therefore, the partial derivative matrix of the dependent variable, for each unit (i.e., area), with respect to explanatory variable k at any given time t is written as:

$$\Gamma_t=\begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1,k}} & \cdots & \frac{\partial E(y_1)}{\partial x_{N,k}} \\ \cdots & \cdots & \cdots \\ \frac{\partial E(y_N)}{\partial x_{1,k}} & \cdots & \frac{\partial E(y_N)}{\partial x_{N,k}} \end{bmatrix}_t=(\boldsymbol{I_N}-\boldsymbol{\rho}\mathbf{W})^{-\mathbf{1}}\boldsymbol{I_N}\boldsymbol{\beta}_k \tag{A21}$$

The matrix of partial derivatives Γ_t is independent of the given time *t*. Therefore, the direct effect (of the *k*-th variable in $\boldsymbol{X}$) is the average of the diagonal elements of the matrix $(\boldsymbol{I_N}-\boldsymbol{\rho}\mathbf{W})^{-\mathbf{1}}\boldsymbol{I_N}\boldsymbol{\beta}_k$, while the indirect effect is the average of the row sums (or equivalently, the average of the column sums) of the off-diagonal elements of the same matrix.

[6] $tr()$ denotes the trace of a matrix.

Table F1. Result of Panel Data Spatial Auto-Regressive Model for Total Contribution

	Dependent variable: $y_{it} = NumContribution_{it}$			
Variables	(1) SAR	(2) SAR	(3) DSAR	(4) DSAR
VC_{it}	0.6258***	0.3999***	0.4406***	0.2560***
	(0.0853)	(0.0843)	(0.0745)	(0.0746)
VC_{it}^2	-0.0062***	-0.0033+	-0.0058***	-0.0032*
	(0.0018)	(0.0018)	(0.0013)	(0.0013)
$VC_{it} \times RushHour_t$		1.1829***		0.9883***
		(0.1813)		(0.1535)
$VC_{it}^2 \times RushHour_t$		-0.0159***		-0.0139***
		(0.0034)		(0.0028)
Wy_t	0.3532***	0.3474***	0.3274***	0.3233***
	(0.0415)	(0.0413)	(0.0429)	(0.0429)
Wy_{t-1}			0.0304	0.0275
			(0.0321)	(0.0315)
y_{t-1}			0.1902***	0.1865***
			(0.0414)	(0.0409)
$Log(UserSpeed_{it})$	-1.1674***	-1.0945***	-0.8002***	-0.7483***
	(0.1253)	(0.1198)	(0.1037)	(0.0990)
$Log(TrafficJamDen_{it})$	2.4504***	2.3360***	2.2076***	2.1177***
	(0.4633)	(0.4428)	(0.3876)	(0.3716)
$Log(AccidentDen_{it})$	3.8202***	3.7672***	2.9864***	2.9633***
	(1.0226)	(1.0033)	(0.7547)	(0.7476)
$Log(VictimDen_{it})$	-0.3842	-0.3648	-0.0714	-0.0625
	(0.6896)	(0.6802)	(0.5833)	(0.5799)
$Temperature_{it}$	-0.0575	-0.0483	-0.0442	-0.0368
	(0.0647)	(0.0640)	(0.0534)	(0.0531)
$Humidity_{it}$	-0.0090	-0.0114	-0.0092	-0.0111
	(0.0148)	(0.0147)	(0.0127)	(0.0126)
$WindSpeed_{it}$	0.0004	0.0045	0.0017	0.0051
	(0.0223)	(0.0219)	(0.0177)	(0.0174)
NTA fixed effects	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes
Borough dummies	No	No	No	No
Date dummies	Yes	Yes	Yes	Yes
Hour-of-day dummies	Yes	Yes	Yes	Yes
Number of Obs.	442,176	442,176	442,176	442,176
Number of NTAs	188	188	188	188
Adjusted R^2 (within)	0.0351	0.0659	0.127	0.142

Note. Robust standard errors in parentheses. Estimates for weather dummies, borough dummies, date, and hour-of-day indicators are omitted from the above results. Time-variant control variables include $Log(UserSpeed_{it})$, $Log(TrafficJamDen_{it})$, $Log(AccidentDen_{it})$, $Log(VictimDen_{it})$, and weather controls. *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table F2. Marginal Effect Estimation of Panel Data Spatial Auto-Regressive Model for Total Contribution

	Dependent variable: $y_{it} = NumContribution_{it}$					
	(1) SAR			(2) SAR		
	Direct	Indirect	Total	Direct	Indirect	Total
VC_{it}	0.6467	0.3209	0.9676	0.4127	0.2001	0.6128
VC_{it}^2	-0.0064	-0.0032	-0.0096	-0.0034	-0.0016	-0.0050
$VC_{it} \times RushHour_t$				1.2208	0.5917	1.8126
$VC_{it}^2 \times RushHour_t$				-0.0164	-0.0080	-0.0244
	(3) DSAR			(4) DSAR		
	Short-Term			Short-Term		
	Direct	Indirect	Total	Direct	Indirect	Total
VC_{it}	0.4529	0.2020	0.6550	0.2630	0.1153	0.3784
VC_{it}^2	-0.0060	-0.0027	-0.0087	-0.0033	-0.0015	-0.0048
$VC_{it} \times RushHour_t$				1.0153	0.4452	1.4605
$VC_{it}^2 \times RushHour_t$				-0.0142	-0.0062	-0.0205
	Long-Term			Long-Term		
	Direct	Indirect	Total	Direct	Indirect	Total
VC_{it}	0.5745	0.4001	0.9746	0.3314	0.2219	0.5533
VC_{it}^2	-0.0076	-0.0053	-0.0129	-0.0042	-0.0028	-0.0070
$VC_{it} \times RushHour_t$				1.2792	0.8568	2.1359
$VC_{it}^2 \times RushHour_t$				-0.0180	-0.0120	-0.0300

References

Anselin, L., Rey, S. J. (2014) Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace, and PySAL. GeoDa Press, Chicago, IL.

Elhorst, J.P. (2014) Spatial Econometrics from Cross-Sectional Data to Spatial Panels. Springer.

Haviland, A., Nagin, D.S., Rosenbaum, P.R. (2007) Combining Propensity Score Matching and Group-Based Trajectory Analysis in An Observational Study. Psychological Methods 12(3): 247-267.

LeSage, J., R. K. Pace. 2009. *Introduction to Spatial Econometrics*. Taylor & Francis, Boca Raton, FL.

Lloyd, C. D. 2010. *Spatial Data Analysis: An Introduction for GIS Users*. Oxford University Press, Oxford, UK.

Vuong, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**(2) 307–333.

Xu, K., Chan, J., Ghose, A., Han, S.P. (2017) Battle of the Channels: The Impact of Tablets on Digital Commerce. Management Science, 63(5): 1271-1656.