Factors Affecting Demand for Shared Bicycles: A Linear model based study

Lu Chen, Micky Sun, Yiting Yu and Zuqi Li
The Department of Economics, University of Southern California

Econ 513: Practice of Econometrics

Professor Manochehr Rashidian

December 9,2022

Introduction

With the rapid development of urbanization, automobile transportation plays an important role in people's daily life. However, the rise of environmental pollution, chaotic traffic and noise continue to become a challenge all around the world and the task is to find a new substitute for motor transportation. According to the data, automobile transport has the biggest impact on urban centers: 71% of climate change, 51% of noise and 95% of atmospheric pollution (Boryaev et al., 2020). Public bike sharing, a more sustainable and flexible option, can be the better solution to these environmental problems and public transportation. For short-distance trips, bike sharing can greatly improve and contribute to the car decrease and transportation circulation (Li & Kamargianni, 2018).

In countries such as the United States, bicycle sharing is encouraged by the government and it's becoming a common sight in urban areas. The world is witnessing the growing demand and investment in this business. The market shows strong resilience and unlimited opportunities. There are multiple factors that jointly affect demand, for example, weather situations and people's life routines. With more bike sharing tech unicorn companies emerging, it's getting more important to investigate factors affecting the demand. This paper investigates two categories of variables, natural and anthropogenic factors, to explore whether these factors have effects on bike sharing demand and what effects they have. The research analyzes the data of rental bike count from 2011 to 2012 in the Capital bikeshare system utilizing linear estimations involving both cross-sectional data analysis and time series analysis, verifying that people tend to rent shared bikes when the weather is warm and clear, echoing with previous literature. The study also finds that people are more likely to rent bikes on workdays and specifically during rush hours, namely 8 am, 5 pm and 6 pm. Hopefully, the learning would be helpful in predicting demand volume and coordinating tech companies' operations.

Literature Review

As Huthaifa, Mohammed and Hesham stated, BSSs (bike-sharing systems) play an important role in urban mobility and eco-friendly mode of traveling. In order to get a better understanding of how different weather conditions and time periods would affect the demand for bike sharing, Huthaifa, Mohammed and Hesham used count models which were based on generalized linear models to predict the number of bikes at specified time periods and weather conditions in each station for San Francisco Bay Area (Ashqar et al., 2019). They concluded that the different time periods and some weather variables are significant determinants (Ashqar et al., 2019). Instead of precipitation, humidity is a significant predictor in Bay Area since the most common rain here is light and moderate rain which has little effect on bike sharing here.

Meanwhile, the time of day has a major impact on demand, meaning that the bike count in each station oscillates for different times such as peak and off-peak time periods (Ashqar et al., 2019).

In a study by Wafic, Mohamed and Khandker, they used the log-linear and lag model regression to estimate and examine the effects of weather, socio-demographic elements, and the use of land and environment on bike share in Toronto. The real-time data of ridership was collected. They used the data sets to perform regressions on the frequency of riders' use on weekdays and weekends, the location of bike-sharing stations, and different routines, combining with the weather variables in the model (El-Assi et al., 2015). As a result, they concluded that higher temperature and lower humidity have a positive influence on the riders' use. Meanwhile, the location of the bike-sharing station is also significant in the estimation. The stations' nearby university campuses, transit spots and popular viewpoints have higher demand (El-Assi et al., 2015).

Data and Methodology

Our purpose is to figure out the important factors affecting the usage of sharing bikes in order to give advice on improving the operation of bike-sharing systems. From the perspective of common sense and literature review, we can categorize variables into natural factors and anthropogenic ones. Natural factors include season, weathersit, temperature, humidity and wind speed. Anthropogenic factors including whether the day is a holiday and whether it is a workday are important factors for customers to decide whether they should rent a bike. Holiday here refers to legal holidays, which do not coincide with weekends. Thus, the qualitative variables holiday and workday can be taken into account at the same time. The following Table 1 shows the definition of variables analyzed.

Table 1 Definition of Variables

Variable	Variable Name	Definition
Usage of sharing bikes	cnt	Count of total sharing bikes including both casual and registered.
Natural factors		
Temperature	$temp$	Normalized temperature. (°C)
Humidity	hum	Normalized humidity. (%)
Wind speed	$windspeed$	Normalized wind speed. The values are divided to the maximum 67.
Weather sit	$weathersit_1$	= 1 if the weather is clear or with few clouds; = 0 otherwise.

	$weathersit_2$	= 1 if the weather is misty or cloudy; = 0 otherwise.
	$weathersit_3$	= 1 if the weather is light rain or light snow; = 0 otherwise.
	$weathersit_4$	= 1 if the weather is severe with heavy rain or heavy snow; = 0 otherwise.
Season	$season_1$	= 1 if it is Winter; = 0 otherwise.
	$season_2$	= 1 if it is Spring; = 0 otherwise.
	$season_3$	= 1 if it is Summer; = 0 otherwise.
	$season_4$	= 1 if it is Fall; = 0 otherwise.
Anthropogenic factors		
Holiday	$holiday$	= 1 if it is legal holiday like Christmas break, Thanksgiving break and so on; = 0 otherwise.
Workday	$workday$	= 1 if it is workday; = 0 if it is weekend.
Hour	$hour_i$	= 1 if it is in the i^{th} hour of the day; = 0 otherwise.

The data used for analysis is a data set sourced from Hadi Fanaee-T(2013). The dependent variable we want to analyze is the count of total rental bikes, and the data are hourly and daily data from the Capital Bikeshare system in Washington DC from 2011-2012. Weather data is from the iweather website. The holiday data are from the government of DC. And from the data set we form, we can obtain the summary of descriptive statistics as Table 2.

Table 2 Summary of Descriptive Statistics

Variables	mean	sd	min	max
temp	0.497	0.193	0.0200	1
atemp	0.476	0.172	0	1
hum	0.627	0.193	0	1
windspeed	0.190	0.122	0	0.851
cnt	189.5	181.4	1	977

Correlation between variables are exhibited in Table 3. Temperature and felt temperature are highly correlated, of correlation 0.9877. Variable atemp is eliminated for the following study to control model efficiency.

Table 3 Summary of variable correlation

	cnt	temp	atemp	hum	windspeed	working day	holiday
cnt	1.0000						
temp	0.4048	1.0000					
atemp	0.4009	0.9877	1.0000				
hum	-0.3229	-0.0699	-0.0519	1.0000			
windspeed	0.0932	-0.0231	-0.0623	-0.2901	1.0000		
workingday	0.0303	0.0554	0.0547	0.0157	-0.0118	1.0000	
holiday	-0.0309	-0.0273	-0.0310	-0.0106	0.0040	-0.2525	1.0000

This research constructs ordinary least square models first to analyze the factors that have significant impact on the count of total rental bikes.

$$cnt_t = \beta_0 + \beta_1 temp_t + \beta_2 hum_t + \beta_3 windspeed_t + \beta_4 holiday + \beta_5 workday + \sum_{i=2}^{4} \alpha_i \, season_i + \sum_{i=2}^{4} \gamma_i \, weathersit_i \quad (1)$$

$$cnt_t = \beta_0 + \beta_1 temp_t + \beta_2 hum_t + \beta_3 windspeed_t + \beta_4 holiday + \beta_5 workday + \sum_{i=2}^{4} \alpha_i\ season_i + \sum_{i=2}^{4} \gamma_i\ weathersit_i + \sum_{i=2}^{24} \delta_i\ hour_i \qquad (appendix\ table\ 1)$$

From the result of this regression attached in appendix table 1, we notice that most coefficients are significant and the coefficients of $hour_i$ have great differences from each other. There exists high demand for rental bikes in some specific time periods. Defining a threshold of 300, we can separate hours in the day into rush hours (including 8, 17, and 18 in the day) and non-rush hours (including other hours). Thus we assign a new dummy $rush = \begin{cases} 1, & if\ it\ is\ 8, 17\ or\ 18\ in\ the\ day. \\ 0, & otherwise. \end{cases}$ Substituting the in equation (1) with , we construct a new regression model.

$$cnt_t = \beta_0 + \beta_1 temp_t + \beta_2 hum_t + \beta_3 windspeed_t + \beta_4 holiday + \beta_5 workday + \beta_6 rush + \sum_{i=2}^{4} \alpha_i\ season_i + \sum_{i=2}^{4} \gamma_i\ weathersit_i \qquad (2)$$

Considering the model is based on time series data and most variables are also highly correlated with time. To isolate the effects of these independent variables on the dependent variable excluding the time effect, we add a variable t to the right hand side of the regression as equation (3) shows. Detrended variables are regressed in (4) to see if we get an appropriate R2 for our model.

$$cnt_t = \beta_0 + \beta_1 temp_t + \beta_2 hum_t + \beta_3 windspeed_t + \beta_4 holiday + \beta_5 workday + \beta_6 rush + \sum_{i=2}^{4} \alpha_i\ season_i + \sum_{i=2}^{4} \gamma_i\ weathersit_i + \theta_1 t \qquad (3)$$

$$cnthat_t = \beta_0 + \beta_1 temp_t + \beta_2 hum_t + \beta_3 windspeed_t + \beta_4 holiday + \beta_5 workday + \beta_6 rush + \sum_{i=2}^{4} \alpha_i\ season_i + \sum_{i=2}^{4} \gamma_i\ weathersit_i + \theta_1 t \qquad (4)$$

The model involves significant heteroskedasticity as shown in Table 4. The following models alleviate heteroskedasticity by utilizing robust OLS estimation.

Table 4 Test statistics for heteroskedasticity

test	Breush-Pagon test	White's test

Test stat (chi2)	1277.32	1107.84
Degree of freedom	1	56
p-value	0.0000	0.0000

Despite the separating time trend, the OLS models have residuals that exhibit autocorrelation. This research further verifies factors' effects in time series models and checks the characteristics of dependent variable volatility.

Before entering variables into time series models, we need to check the stationarity. Traditionally, differencing, taking log or taking the square root, and taking the moving average is used to smooth the variables. This research utilizes the moving average of 72 lags to satisfy data with stationarity since almost half of the variables of interest are dummy variables, which would be eliminated by differencing. The unit-root test statistics are shown in the following table.

Table 6 Dicky-Fuller test statistics of variables

Variables	macnt	matemp	mahum	mawindspeed
Z(t)	-2.336	-0.585	-2.239	-3.916
p-value	0.1606	0.8744	0.1925	0.0019

Note: macnt, matemp, mahum, mawindspeed refer to the moving average of 72 lags of rented shared bike count, temperature, humidity and wind speed.

This research utilizes autoregressive moving average (ARMA) for level modeling and autoregressive conditional heteroskedasticity (ARCH) for volatility modeling. For ARMA model, suitable parameters are specified for ARMA (p, q), where p is the dimension of the autoregressive component, and q is the dimension of the moving average component of the model (Ives et al., 2010). The ARCH model was presented by Engle (1982) suggesting that the variance of the residuals at the time t depends on the squared error terms from past periods. The time series models of order (1, 0) and ARCH (1) are specified considering both simplicity and economic meaning (appendix table 2) for the count of shared bike demand as below:

$$macnt_t = \sum_{i=2}^{4} \alpha_i \, season_i + \beta_1 holiday_t + \beta_2 workingday_t + \sum_{i=2}^{4} \gamma_i \, weathersit_i + \beta_3 matemp_t + \beta_4 mahum_t + \beta_5 mawindspeed_t + \beta_6 rush_t + macnt_{t-1} + c + u_t \quad (5)$$

In the equation above, c refers to a constant, $macnt_{t-1}$is the moving average of count in the previous period, u_t is the error term at time t, and the rest are contemporary variables.

$$macnt_t = \sum_{i=2}^{4} \alpha_i \, season_i + \beta_1 holiday_t + \beta_2 workingday_t + \sum_{i=2}^{4} \gamma_i \, weathersit_i + \beta_3 matemp_t + \beta_4 mahum_t + \beta_5 mawindspeed_t + \beta_6 rush_t + macnt_{t-1} + c + u_t + \alpha {u_{t-1}}^2 \quad (8)$$

ARCH (1) model differs from ARMA(1,0) in not having the moving average of count in the previous period plus an additional variance from the previous period, denoted by ${u_{t-1}}^2$.

Empirical Results

As shown in the table, both natural factors and anthropogenic factors have significant effects on bike-sharing demand. Natural factors, weather situations, temperature, humidity and wind speed affect the count of shared bikes rented. People tend to rent bikes in misty weather, presumably because misty weather is of best riding comfortability. The increase in temperature significantly increases the number of bikes rented. The increase is up to 400.718 bikes when the temperature varies from the lowest to the highest. Washington DC people rent significantly fewer bikes when the weather gets more humid. As the humidity increases, people decrease their rent by up to 240.486. Wind speed can decrease the number of bikes rented by 17.997.

On the other hand, anthropogenic factors have significant effects on the total count of shred bikes rented as well. On holiday, the number of bikes rented on average decreases by 30.517. People tend to rent 237.533 more bikes on average during rush hours, which is 8 am in the morning and 5 pm and 6 pm in the evening.

Table 7 Ordinary least squares estimation results

	(1)	(2)	(3)	(4)

VARIABLES	cnt	cnt	cnt	cnthat
spring	-0.987	5.655*	2.336	1.529
	(4.160)	(3.343)	(3.200)	(3.216)
summer	-43.404***	-33.006***	-50.896***	-55.248***
	(5.317)	(4.704)	(4.448)	(4.459)
fall	57.090***	57.608***	16.377***	6.348**
	(3.654)	(2.889)	(2.892)	(2.802)
holiday	-27.610***	-27.288***	-30.517***	-31.303***
	(7.212)	(5.947)	(5.913)	(5.984)
workingday	1.727	2.146	2.785	2.940
	(2.595)	(2.386)	(2.322)	(2.332)
mist	17.023***	14.226***	11.503***	10.841***
	(2.842)	(2.369)	(2.292)	(2.300)
light	-3.437	-16.373***	-17.042***	-17.205***
	(4.743)	(3.812)	(3.695)	(3.718)
heavy	104.820	39.757	33.701	32.228
	(88.755)	(32.651)	(43.500)	(46.819)
temp	454.794***	421.955***	400.718***	395.552***
	(9.963)	(8.621)	(8.216)	(8.169)
hum	-286.652***	-252.248***	-240.486***	-237.624***
	(7.221)	(6.352)	(5.986)	(5.946)
windspeed	31.653***	7.057	17.997**	20.658**
	(10.202)	(8.570)	(8.216)	(8.233)

rush		236.608***	237.533***	237.758***
		(4.358)	(4.066)	(4.015)
t			0.008***	
			(0.000)	
Constant	130.220***	96.891***	43.457***	-71.542***
	(6.775)	(5.592)	(5.467)	(5.327)
Observations	17,379	17,379	17,379	17,379
R-squared	0.283	0.468	0.509	0.465

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The time series model uses the moving average of 72 lags of variables, smoothing out the strikes resulting from anthropogenic factors involving rush hours. However, natural factors like weather conditions, temperature, humidity and wind speed still have significant effects on the amount of bikes rented. The directions of the effect remain the same. Besides, the model exhibits a significant ARCH effect implying the demand involves clustering volatility.

Table 8 Time series model estimation results

	(5)	(6)	(7)	(8)	(9)
VARIABLES	macnt	ARMA	sigma	macnt	ARCH
spring	-0.175			41.400***	
	(0.484)			(0.244)	
summer	-0.684			0.572*	
	(0.511)			(0.327)	
fall	0.244			45.133***	
	(0.391)			(0.173)	
holiday	-0.006			-2.094***	
	(0.109)			(0.298)	

workingday	-0.462***			-2.317***	
	(0.051)			(0.102)	
mist	0.130***			-0.531***	
	(0.049)			(0.104)	
light	0.269***			-0.490***	
	(0.079)			(0.138)	
heavy	0.034			14.358***	
	(0.624)			(3.967)	
matemp	227.628***			290.973***	
	(11.101)			(0.660)	
mahum	-128.616***			-166.868***	
	(5.234)			(0.584)	
mawindspeed	-47.152***			-237.394***	
	(6.914)			(0.917)	
rush	-0.020			-0.129*	
	(0.062)			(0.077)	
L.ar		0.999***			
		(0.000)			
L.arch					1.007***
					(0.047)
Constant	154.680***		2.100***	151.809***	2.176***
	(25.535)		(0.018)	(0.446)	(0.109)
Observations	17,378	17,378	17,378	17,378	17,378

Note: standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Conclusion

This paper researches factors affecting the bike-sharing demand and the specific effect each factor has. Both natural and anthropogenic factors investigated above have significant impacts on shared bike demand. The Washington DC bike sharing data shows that people rent shared bikes when the weather is warm and clear which obtained the same result from Wafic's study. People are more likely to rent bikes on workdays and specifically during rush hours, namely 8 am, 5 pm and 6 pm. The result has limitations as well. During the data collection period, the bike-sharing company did not adjust the unit price. The research is restrained from adding price into the model. Future studies can be done to incorporate more anthropogenic factors including prices, and sustainable policies to better model the amount of shared bikes demanded. Moreover, autoregressive conditional heteroskedasticity can also be added to the time series model and would help better predict the amount demanded in the future.

References

Ashqar, H. I., Elhenawy, M., & Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. *Case Studies on Transport Policy*, *7*(2), 261–268.

Boryaev, A., Malygin, I., & Marusin, A. (2020). Areas of focus in ensuring the environmental safety of Motor Transport. *Transportation Research Procedia*, *50*, 68–76.

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of theEconometric Society*, 987-1007.

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2015, December 16). Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto - transportation. SpringerLink. Retrieved December 8, 2022, from https://link.springer.com/article/10.1007/s11116-015-9669-z#Sec7

Fanaee-T, Hadi, and Gama, Joao, 'Event labeling combining ensemble detectors and background knowledge', Progress in Artificial Intelligence (2013): pp. 1-15.

Ives, A. R., Abbott, K. C., & Ziebarth, N. L. (2010). Analysis of ecological time series with ARMA (p, q) models. *Ecology*, 91(3), 858-871.

Li, W., & Kamargianni, M. (2018, March 27). *Providing quantified evidence to policy makers for promoting bike-sharing in heavily air-polluted cities: A mode choice model and policy simulation for Taiyuan-China*. Transportation Research Part A: Policy and Practice. Retrieved December 8, 2022, from https://www.sciencedirect.com/science/article/pii/S0965856417301003

Appendices

Appendix Table 1

OLS Estimation with hour dummies

	(1)	(2)	(3)
VARIABLES	cnt	cnt	cnthat
spring	35.502***	33.430***	33.131***
	(2.909)	(2.665)	(2.664)
summer	15.356***	-1.546	-3.988
	(4.051)	(3.679)	(3.672)
fall	62.552***	18.038***	11.608***
	(2.483)	(2.456)	(2.350)
holiday	-25.291***	-28.683***	-29.173***
	(5.402)	(5.426)	(5.486)
workingday	4.210**	5.015**	5.131**
	(2.092)	(1.994)	(2.000)
mist	-6.130***	-10.000***	-10.559***
	(1.946)	(1.828)	(1.829)
light	-61.529***	-64.427***	-64.846***
	(3.393)	(3.296)	(3.318)
heavy	-38.737	-50.252	-51.915
	(64.771)	(58.502)	(58.026)
temp	276.225***	245.607***	241.184***
	(7.862)	(7.169)	(7.120)
hum	-101.114***	-80.814***	-77.882***
	(5.796)	(5.373)	(5.336)

windspeed	-45.104***	-35.144***	-33.705***
	(7.499)	(6.983)	(6.988)
1.hr	-16.881***	-17.365***	-17.435***
	(2.949)	(3.429)	(3.595)
2.hr	-25.015***	-26.355***	-26.549***
	(2.945)	(3.475)	(3.646)
3.hr	-35.070***	-36.996***	-37.274***
	(2.863)	(3.436)	(3.613)
4.hr	-37.057***	-39.922***	-40.335***
	(2.857)	(3.412)	(3.588)
5.hr	-20.882***	-23.250***	-23.592***
	(2.770)	(3.288)	(3.463)
6.hr	37.569***	35.633***	35.354***
	(2.937)	(3.212)	(3.361)
7.hr	171.986***	170.607***	170.408***
	(5.721)	(5.527)	(5.568)
8.hr	311.367***	310.946***	310.885***
	(8.319)	(7.943)	(7.935)
9.hr	161.969***	162.953***	163.096***
	(3.579)	(3.260)	(3.322)
10.hr	105.409***	107.856***	108.209***
	(3.933)	(3.824)	(3.902)
11.hr	129.040***	132.998***	133.570***
	(4.661)	(4.464)	(4.512)
12.hr	166.893***	172.093***	172.844***

	(5.157)	(4.869)	(4.893)
13.hr	160.968***	166.901***	167.758***
	(5.305)	(5.033)	(5.056)
14.hr	144.546***	150.913***	151.833***
	(5.336)	(5.111)	(5.141)
15.hr	153.770***	160.243***	161.178***
	(5.232)	(4.957)	(4.983)
16.hr	216.096***	222.305***	223.202***
	(4.833)	(4.390)	(4.400)
17.hr	370.550***	376.110***	376.913***
	(7.418)	(6.967)	(6.936)
18.hr	339.716***	344.421***	345.101***
	(7.237)	(6.819)	(6.795)
19.hr	232.575***	236.110***	236.621***
	(4.979)	(4.649)	(4.670)
20.hr	154.238***	156.777***	157.143***
	(3.712)	(3.543)	(3.616)
21.hr	105.810***	107.517***	107.764***
	(3.008)	(3.030)	(3.149)
22.hr	69.599***	70.614***	70.760***
	(2.705)	(2.905)	(3.053)
23.hr	31.360***	31.840***	31.909***
	(2.646)	(2.985)	(3.148)
t		0.009***	
		(0.000)	

Constant	-22.424***	-83.593***	-194.430***
	(5.237)	(5.253)	(5.116)
Observations	17,379	17,379	17,379
R-squared	0.627	0.675	0.647

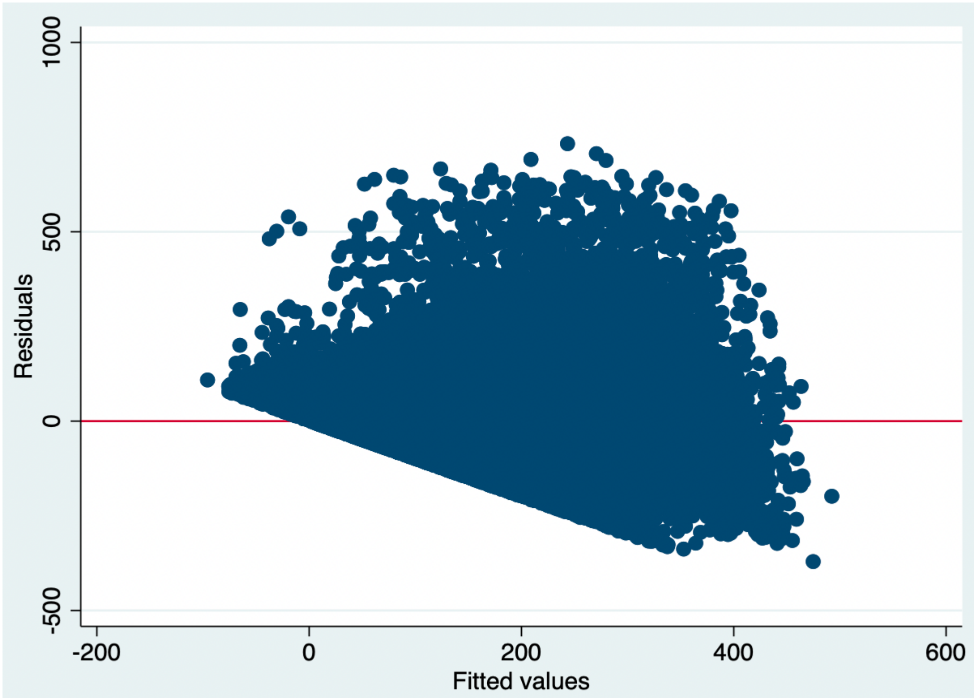
Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 2

	(1)
VARIABLES	u
L.u	0.838***
	(0.008)
L2.u	-0.107***
	(0.010)
L3.u	-0.022***
	(0.008)
Constant	-0.003
	(0.661)
Observations	17,376
R-squared	0.566

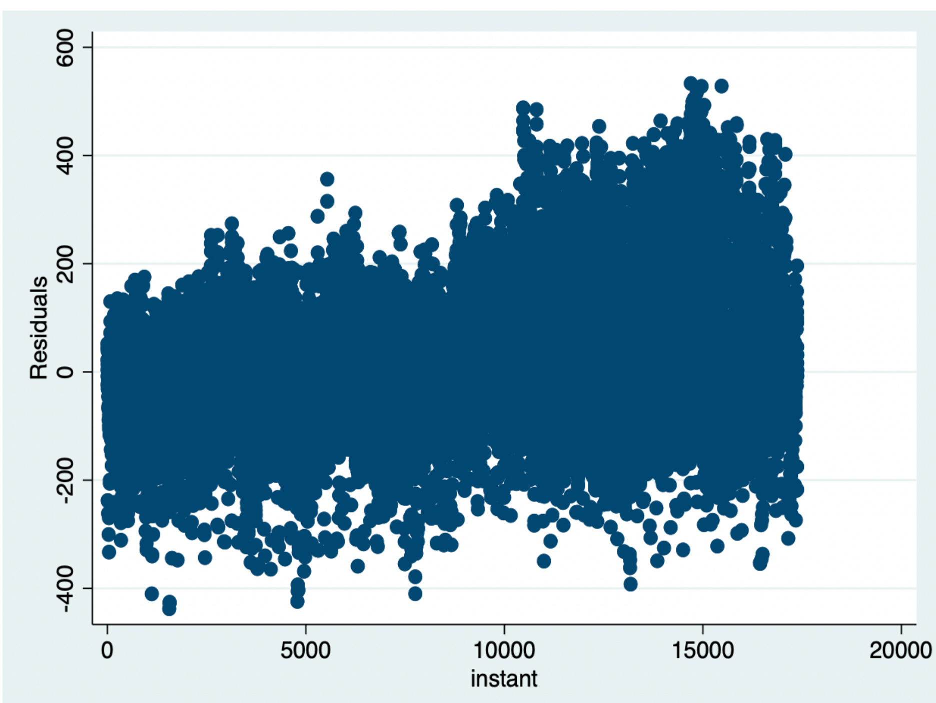
Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Fig 1 Residual visualization



Note: instant refers to time.

Appendix Fig 2 Autocorrelation Visualization



Note: instant refers to time.