Examination of the Broken Windows Theory Using User-Generated Data

The following report uses artificially generated data for all variables and does not represent a real neighborhood. This report functions as a thought experiment to showcase an understanding of regression analysis and related topics. Conclusions formed herein should not be taken to reflect real world policy or decisions.

Background:

The Broken Windows Theory holds that visible signs of crime (e.g. broken windows or graffiti) lead to further crime due to the appearance of an environment that condones such behavior. By correcting minor infractions and creating a lawful atmosphere, crime rates in general should decrease. This theory was first presented by James Q. Wilson and George L. Kelling in 1982 and has been the subject of much debate. In the 1990s, this was a driving factor in New York City's controversial "stop and frisk" policy. Areas with a lot of visible crime received further police presence. The purpose of this report is to test the validity of this theory and whether it is relevant for public policy decisions.

Data:

Two datasets are used for the study. The first dataset tracks the presence of littering and graffiti in 200 neighborhoods in urban, suburban, and rural areas. If littering or graffiti is present, the corresponding variable takes a value of 1 and if not, a value of 0. Likewise, population density is also tracked as a binary variable, a proxy for urban and rural areas. Finally, average household income (in thousands) is tracked. Combined with population density, this can help differentiate rural and suburban areas.

```
regress litter graffiti
```

Source	SS	df	MS	Number of obs	=	200
				F(1, 198)	=	7.69
Model	1.17402813	1	1.17402813	Prob > F	=	0.0061
Residual	30.2209719	198	.152631171	R-squared	=	0.0374
				Adj R-squared	=	0.0325
Total	31.395	199	.157763819	Root MSE	=	.39068

litter	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
graffiti	.1549872	.0558828	2.77	0.006	.0447853	.2651891
_cons	.7391304	.0364311	20.29	0.000	.6672876	.8109733

Analysis:

Initial analysis assumes a simple linear regression between the presence of graffiti and the act of littering. With a t-value of 2.77, there appears to be a statistically significant relationship between the two variables however the R-squared value is notably low, suggesting that this model is too simple to properly depict the relation.

```
. regress litter graffiti income high_density
```

Source	SS	df	MS	Number of obs	=	200
				F(3, 196)	=	74.24
Model	16.6989823	3	5.56632744	Prob > F	=	0.0000
Residual	14.6960177	196	.074979682	R-squared	=	0.5319
				Adj R-squared	=	0.5247
Total	31.395	199	.157763819	Root MSE	=	.27382

litter	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
graffiti	.4135519	.0784364	5.27	0.000	.2588642	.5682397
income	.0171332	.0012321	13.91	0.000	.0147034	.0195631
high_density	-.1190648	.0749989	-1.59	0.114	-.2669731	.0288436
_cons	-.2051047	.0710219	-2.89	0.004	-.3451698	-.0650395

The analysis is expanded above to include the binary control variables “income” and “high_density”. When these are accounted for, the t-value of the correlation coefficient between graffiti and litter

increases to 5.27, clearly significant at the 1% level. Similarly, the R-squared value has increased to .53, suggesting that the new model and correlation coefficients more correctly identify the relation. However, there is still an issue of potential simultaneity bias between litter and graffiti. That is, litter may influence graffiti and graffiti may influence litter at the same time. Since both are minor crimes and give rise to an atmosphere of lawlessness, its not clear which, if either, causes the other.

```
. summarize litter high_density income if graffiti == 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
litter	115	.7391304	.4410306	0	1
high_density	115	.1130435	.3180317	0	1
income	115	55.89686	17.77671	23.85765	79.76493

```
. summarize litter high_density income if graffiti == 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
litter	85	.8941176	.309513	0	1
high_density	85	.9529412	.2130215	0	1
income	85	46.64219	15.74409	20.01258	76.27644

Looking at summary statistics for the areas with and without graffiti, its clear that the areas are fundamentally different. Average income is nearly $9000 higher, and likelihood of high population density is 80% lower in areas without graffiti while the occurrence of litter is not particularly different. While having graffiti present appears to have some effect on litter, it seems more likely that population density and average household income are more effective determinants of graffiti and litter. Unobserved variables like the police-to-resident ratio are likely important here. Other studies like the 1985 “Relationship Between Police Presence and Crime Deterrence” by J. Rouse suggest that increased police presence (not necessarily police-to-resident ratio) is highly deterministic with crime rates.

```
. regress litter graffiti high_density income, vce(robust)

Linear regression                               Number of obs     =        200
                                                F(3, 196)         =      40.16
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4683
                                                Root MSE          =     .27668

------------------------------------------------------------------------------
             |               Robust
      litter |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    graffiti |   .1126774   .0396909     2.84   0.005     .0344013    .1909534
high_density |   .0170671   .0391729     0.44   0.664    -.0601874    .0943217
      income |   .0146544   .0013459    10.89   0.000     .0120001    .0173088
       _cons |   .0097763   .0917305     0.11   0.915    -.1711291    .1906817
------------------------------------------------------------------------------
```

To combat the issue of unobserved variables or simultaneity bias, graffiti and litter were first cleaned from a neighborhood, then new graffiti was placed in several neighborhoods at random. This data was collected in a new dataset and regression analysis was again conducted. The above data shows that graffiti still has a positive correlation with litter, however the size of the effect is diminished from the first dataset. It is still statistically significant, however the size of the coefficient between graffiti and litter has decreased to just 0.11 from the 0.43 of the earlier dataset. Notably, high population density is not statistically significant, suggesting that graffiti placed in urban or suburban areas had similar effects assuming all else is equal.

Conclusion:

While there does appear to be a positive correlation between litter and graffiti, this does not necessarily validate the broken window theory for use in public policy The theory holds that the presence of observably broken laws leads to further lawbreaking. This study shows only a small but statistically significant increase in littering with the presence of observable broken laws which only accounts for 46%

of the observed variance. Furthermore, the presence of both graffiti and litter is shown to increase in low-income, high population areas. It is more likely that a third variable relating to the economic status of the neighborhood is more relevant in determining crime rates.