The Association of Aircraft Operations and Air Quality

Reggie Arkell
Southern New Hampshire University
ECO 625: Applied Econometrics II
9-1 Final Project: Econometric Analysis
August 20, 2017

Description

Community advocates often seek new and expanded airports to achieve economic development goals. It is well known that populations residing near airports experience negative externalities from aircraft noise and associated air emissions, including comparable impacts from affiliated roadway traffic. The impacts often tend to increase over time in areas with substantive population growth and increasing travel demand. Less evident is the extent that these negative externalities, particularly the concentrated air pollution, impact regions as a whole. The study seeks to quantify these effects. The research is relevant to economics as general perceptions are that increasing passenger airport capacity provides significant benefits in terms of employment, income and gross domestic product (GDP) that outweigh social costs. The target audience of this research is the general public, urban/regional planners, and political decisionmakers. The key finding is that increases in flights and flight delays have a statistically significant positive association with higher air pollution in metropolitan areas as measured by the Air Quality Index (AQI), albeit, the increases are very small.

Literature Review

According to a study by the Partnership for AiR Transportation Noise and Emissions Reduction, aircraft landing and take-off operations in the U.S. during 2002 contributed less than 1 percent of total emissions for 5 criteria pollutants in 118 nonattainment areas with at least 1 commercial airport. However, the high-end ranges for these were as follows: carbon monoxide (CO, 4.36 percent), nitrous oxide (NO_X, 10.93 percent), volatile organic compounds (VOCs, 5.03 percent), oxides of sulfur (SO_X, 6.91 percent) and particulate matter 2.5 ($PM_{2.5}$, 2.57 percent) (Ratliff et al., 2009).

Lin et al. performed multivariate logistic cross-sectional regression in a study of 3 airports (1 large, 1 medium, 1 small) in the state of New York regarding respiratory health for residents within close proximity of these facilities for the period 1995-2000. Detailed hospital admissions data was obtained for patients (age, sex, race address, etc.) from the New York State Department of Health via the Statewide Planning and Research Cooperative System (SPARCS). Controls included wind speed and direction. Data was not obtained on airport or aircraft operations in terms of numbers of flights or associated congestion/delay. The authors found that for 2 of the airports, persons living within 5 miles of these facilities are associated with increased hospital admission rates for respiratory ailments compared to those living farther away. The third airport did not have any such relationship (Lin et al., 2008).

In 2008, Lu studied operations at Taiwan Taoyuan International Airport to identify economic costs and benefits. Social costs were calculated for 6 criteria pollutants attributed to emissions estimates for the different aircraft based upon operating phases of landing, take-off, and 30 minutes of cruise. The total annual social cost of aircraft emissions at the airport is about €86.7 million (Euros) while aircraft noise is about €12 million. Total gross economic benefit, not accounting for opportunity costs, is about €913 million. Thus, while aircraft emissions and noise have substantive social costs, they are a relatively small proportion (~10 percent) of the economic benefits (Lu, 2011).

Schlenker and Walker studied the impacts of operations, particularly runway congestion, at the 12 largest California airports (by passenger count) on local air pollution and hospital admissions for the years 2005-2007. A substantial proportion of airport emissions are generated by airplanes overall and when idling. As with most transport modes, jet engines from aircraft generate carbon dioxide (CO_2), CO, NO_X, SO_X and VOCs among others. These emissions levels

vary by aircraft phase of movement, however, CO and NO_X levels are highest during the idle/taxi stages. Data from the Bureau of Transportation Statistics (BTS) shows that from 1975 to 2007 airplane taxi time increased by 23 percent. This trend is due to the increased use of "hub and spoke" operations which benefit travelers and airports while requiring planes to leave gates and enter waiting queues rather than staying longer at terminals. The largest source of CO and the third largest source of NO_X in the state of California is Los Angeles International Airport (LAX) (Schlenker & Walker, 2012).

Schlenker & Walker use least squares regression models to measure congestion by the amount of time taken for planes to taxi from terminals to runways per BTS data. Controlling variables are weather (temperature, precipitation, wind speed/direction (hourly data NOAA weather stations), scheduling, and holidays. Emissions data is from California Air Resource Board (CARB) pollution monitors and consist of averages for regional monitors based upon hourly/daily postings for CO, nitrogen dioxide (NO_2) and ozone (O_3). Hospital admissions and emergency room data is from the California Emergency Department & Ambulatory Surgery and includes patient age, zip code of residence, primary and secondary diagnosis. The authors found that a change in standard deviation of 1 in LAX airport congestion is associated with a 0.32 standard deviation increase in CO nearby which fades by distance. The mean effect is 0.23 standard deviations in zip codes within about 6.2 miles of the airport. This equates to hospitalization costs of $1 million for the 6 million people living within such distance of an airport. These consist of increases in heart problems by 17 percent, overall respiratory ailments by 18 percent, and asthma counts by 30 percent. No associations were found relative to non-respiratory ailments. Thus, the externalities of airport congestion extend substantially beyond

diminished travel time even when they are within U.S. Environmental Protection Agency (EPA) limits (Schlenker & Walker, 2012).

Shirmohammadi et al. studied particle number (PN), black carbon (BC) and $PM_{2.5}$ levels nearby and within 150 miles downwind of LAX Airport monitors at the I-110, I-405 and I-105 freeways within the airport impact zones during a 3-month period in mid-2016. The research found that respective PN, BC and $PM_{2.5}$ levels were about 11, 2.5, and 1.4 times higher at data gathering points closest to the airport as opposed to the expressways. Nevertheless, the findings for PN were a 4-times decrease within about the last 10 years, likely attributable to engine design changes, evolving regulations and altered fuel formulations. The study also finds that air emissions rates in general attributable to LAX Airport are about 10 percent of the totals within Los Angeles County (Shirmohammadi et al., 2017).

Currie and Walker studied the effects of surface traffic congestion on prematurity and birth weight for mothers residing within 10 km of expressway toll plazas in New Jersey and Pennsylvania both before and after the introduction of electronic fare collection (E-Z Pass) that eliminates traveler queuing and manual payment systems. Using least squares regression, the authors find that initiation of E-Z Pass is associated with mothers within 2 km of a toll plaza experiencing reductions of 10.8 percent in prematurity and 11.8 percent in low birth weight compared to mothers residing within 2-10 km. Their findings are relatively consistent with similar studies of both air pollution in general and surface transport congestion impacts with fetal health (Currie & Walker, 2011).

Hypothesis

The null hypothesis to be tested, or $H_0: B_j = 0$, is that an independent variable (IV) consisting of the sum of number of flights and flight delays per annum (FLTS+FLTDEL) does

not have a relationship with air quality as measured by the dependent variable (DV) of median annual air quality index (AQI), and the slope of the population regression is 0. The alternative hypothesis is $H_1: B_j \neq 0$, meaning that FLTS+FLTDEL does have a relationship with AQI, and the slope of the population regression is not 0.

Data

The study focuses on 117 of the largest metropolitan statistical areas (MSAs) for a 5-year period from 2010 through 2014 using pooled cross-sectional panel data. The inclusion of time series data is advantageous as it accounts for changes over several years and presumably improves accuracy of the developed model compared to a conventional single period cross-sectional analysis. The research considered other IVs or controlling variables including those listed in *Figure 1* below in conjunction with FLTS+FLTDEL to identify associations with AQI.

Figure 1 – Tested Independent Variables and Data Sources

Independent Variables	**Source/Description**	**Acronym**
Energy Proportion/Coal Plants[1]	Brookings Institute	COAL
Coastal Plains and Lowlands[2]	U.S. Geological Survey	CPLL
Northeastern U.S.[2]	U.S. Geological Survey	NE
Mountains[2]	U.S. Geological Survey	MTNS
Annual Normal Temperature[1]	National Oceanic and Atmospheric Admin.	TEMP
Annual Normal Precipitation[1]	National Oceanic and Atmospheric Admin.	RAIN
Weighted Pop. Density	U.S. Census	WPD
Population	U.S. Census	POP
Gross Domestic Product (GDP)	Bureau of Economic Analysis	GDP
Per Capita GDP	Bureau of Economic Analysis	PCGDP
Manufacturing Employment	U.S. Census	MANUF
Flights	Federal Aviation Administration	FLTS
Flight Delays	Federal Aviation Administration	FLTDEL
Enplanements	Federal Aviation Administration	PASS
Urban Avg. Traf./Fwy. Lane Mile	FHWA 2014 Highway Stats. (HM-72)	ADTFL

1. Categorical; 2. Bivariate

Median annual AQI is derived from the daily measures of 6 primary air pollutants, i.e. the pollutant with the highest AQI on a given day. A higher numeric value equates to poorer air

quality. FLTS is the number of annual commercial aircraft operations. FLTDEL consists of departures 15 or more minutes after the scheduled times.

Data Limitations

The data gathered is predominantly at the MSA level. Categorical variables are used for COAL, MTNS, TEMP, RAIN, as data was not or could not be differentiated by year. This is a concern as weather-related metrics can be expected to vary from year to year. The same was initially done with MANUF, however, annual state manufacturing employment data was used to estimate annual numbers for MSAs apart from the 2012 Economic Census direct data. WPD was calculated using available 2000 and 2010 data to continue past annual trends. POP and ADTFL are estimates by federal agencies which are subject to error due to lack of precision. FLTDEL does not account for the actual time or minutes of delay which intuitively would increase emissions. AQI was chosen as the air quality DV as there were more limitations on data availability from EPA for other criteria pollutants. Consequently, the number of MSAs analyzed for AQI is 117 over 5 years which equates to 585 and then reduced by 1 year to 468 observations to address authcorrelation due to the time series nature of the data.

Empirical Approach

The 5 years of data has both cross-sectional and time-series components which brings into question if ordinary least squares (OLS or linear regression) analysis is appropriate. The proposal then is to use panel data in a pooled cross-section. However, this structure is likely problematic due to the associations of data between observations for the same areas over successive years. Thus, the configuration of the error term changes over time. This is a violation of the M5 best linear unbiased estimator (BLUE) assumption that there is no correlation between

the error terms and IVs. The error term u_{it} then has a time-invariant element identified as α_{it} and a time-variant element known as ε_{it} that fluctuates across observations and periods (Hilmer & Hilmer, 2014, p. 5-11).

Initially, correlation and covariance matrices are developed to identify one-on-one associations amongst the variables (*Appendix 1 and 2*). As expected, several IVs are highly associated with each other which is indicative of possible multicollinearity if included in the same model. Stepwise regression is used to build the model to improve robustness as measured by R^2 and F values in addition to IV p values and t scores based on a 0.05 level of confidence. The stepwise sequencing began by running regression on AQI against ADTFL, in addition to a combination of FLTS and FLTDEL (FLTS+FLTDEL) as they are highly correlated. FLTDEL and ADTFL are measures of congestion in terms of aircraft and surface vehicles operating at slow ground speeds or while stationary, which is commonly known to significantly elevate emissions (Barth & Boriboonsomsin, 2009)(Balakrishnan & Hansman. 2010). MTNS and RAIN were found to improve robustness and are retained in the model. COAL and TEMP were also included in the initial model due to statistical significance with the DV. POP, GDP and MANUF all have highly positive correlations with AQI, and with each other, while WPD has a negative association. POP and GDP are very similar in their associations with other variables. Thus, the decision was made to drop GDP while combining the POP+MANUF and FLTS+FLTD as a single IV: (POP+MANUF)*(FLTS+FLTDEL). The initial resulting linear equation is:

$$(AQI)_t = \beta_0 + \beta_1 COAL_t + \beta_2 TEMP_t + \beta_3 MTNS_t + \beta_4 RAIN_t + \beta_5 ADTFL_t + \beta_6((POP + MANUF) * (FLTS + FLTSD))_t + e_t.$$

A histogram was made of the AQI data to reveal a relative normal distribution with some skewing to the right (*Appendix 3*). Nevertheless, there is some concern with the DV of AQI as it

is count data, non-continuous, discrete, and developed by counting rather than ranking. Therefore, OLS may be limited in that it does not consider the non-continuous, non-negative format of the data. Consequently, parameter predictions can be biased, inefficient and unreliable (Hilmer & Hilmer, 2014, pp. 243-7). A Poisson Model was performed/considered, however, it was dismissed as descriptive statistics showed that the DV mean and variance were substantively different, i.e. equi-dispersion did not exist. A Negative Binomial Model was also performed and showed statistical significance between each of the IVs with the DV. This test, in part, confirmed via the likelihood ratio test that alpha is significantly different from zero which is further confirmation that the Poisson distribution is inappropriate. A printout of the Stata Negative Binomial Model run is in *Appendix 4a and 4b*.

An MWD test was then performed to determine whether a linear model or a log-log model provided the best specification of the AQI function:

H_o: Linear Model: Y is a linear function of the X's

H_1: Log-log Model: Log Y is a linear function of the X's or a log of the X's

The MWD test was performed by estimating the linear model and obtaining the estimated Y values, i.e. $\hat{Y}_i$. The log-log model was then estimated to obtain the Y_i values or $\widehat{logY_i}$. Z_{1i} was obtained by $Z_{1i} = \hat{Y}_i - \widehat{logY_i}$. Y was then regressed on the X's and Z_{1i}. Z_{1i} was found to be statistically significant as measured by the standard t test, thus, H_0 was rejected. Z_{2i} was then calculated via Z_{2i} = antilog $(\widehat{lnY_i}) - \hat{Y}_i$. Y was then regressed on the logs of X's and Z_{2i}. The result was that Z_{2i} is not statistically significant, therefore, H_1 is accepted and the log-log version is the more appropriate model (Hilmer & Hilmer, 2014, p. 284). The decision was made to retain the log-log regression for the choice of model as a result of the MWD test. Additionally, the log-log version provides the opportunity to predict a more meaningful change in FLTS and

FLTDEL from the (POP+MANUF)*(FLTS+FLTDEL) IV due to the inherent percent change versus linear format compared to the unit change structure. Nevertheless, predictability is limited by the combined IV.

Results and Robustness

A Durbin-Watson test is run on the aforementioned initial linear model: $d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \rightarrow \frac{27108.6129}{47029.8624} = 0.57641276$. Since *d* is >0 and <2 and closer to 0 than 2, there is positive autocorrelation. There is no lower or upper bound metrics from the Durbin-Watson significance tables as n>200. The following fixed-effects techniques were used in an attempt to reduce the autocorrelation: dummy variable, demeaning, and first-differencing. The dummy variable method resulted in no improvement in the Durbin-Watson test statistic: $d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \rightarrow \frac{24213.1021}{45620.1887} = 0.5307541$. However, there still appears to be positive autocorrelation or the interpretation is at least inconclusive. The demeaning and first-differencing methods were unsuccessful as they resulted in removing all statistically significant relationships in the model. The coefficient of autocorrelation or ρ (lies between -1 and 1) is calculated as follows: $\rho = \left(1 - \frac{d}{2}\right) \rightarrow \left(1 - \frac{0.57641276}{2}\right) = 0.711793619$. ρ is part of the mechanism $u_t = \rho u_{t-1} + v_t$ which states that the error or disturbance value at time *t* relies upon in part its value at time period (*t*-1), the arbitrary term (v_2), and the degree of reliance on the previous value as measured by ρ (rho)(Gujarti & Porter, 2010, p. 157).

The model was then transformed to log-log with exception of the dummy variable MTNS and categorical variable RAIN. COAL and TEMP were no longer statistically significant and were removed from the model. The model is then transformed via generalized difference equations which consisted of lagging the log variables, differencing them, multiplying by ρ, and

subtracting the result from the log variables. Thus, the lost first observation year of 2012 is eliminated. The resulting Durbin-Watson statistic is now $d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \rightarrow \frac{11830.3345}{6737.87787} = 1.75579534$ which is a substantive improvement as it is closer to 2. The resulting equation is in the format of:

- $log(Y_t - pY_{t-1}) = B_1(1-p) + B_2X_{2t} + B_3X_{3t} + B_4X_{4t} + logB_5(X_t - pX_{t-1}) + logB_6(X_t - pX_{t-1}) + v_t$ or
- $Y_t^* = B_1^* + B_2X_t^* + B_3X_t^* + B_4X_t^* + B_5X_t^* + B_6X_t^* + v_t$ or
- $\log(AQI)_t^* = 0.414796_1^* + 0.056667_2(MTNS)_t - 0.024777_2(RAIN)_t + 0.074764_3\log(ADTFL)_t^* + 0.043857_4\log(POP(FLTS + FLTD))_t^*$

The Stata regression summary and results of a Ramsey Regression Error Specification Test (RESET) are depicted in *Figure 2* below. R^2 or the proportion of AQI explained by the model is 0.22. The model's *F* is acceptable at 33.5 and the significance *F statistic* or *P-value* (Prob > F) is statistically significant as it is well below the 0.05 threshold. ADTFL is just on the cusp of statistical significance at the 0.10 level. All other IVs are statistically significant as *p-values* are below the 0.05 level and *t Stat* numbers are above the 1.960 threshold. Thus, when controlling for all other factors, predicted associations are as follows: areas with mountains/valleys (MTNS) and AQI increases of 5.66 percent; annual rainfall reductions (RAIN) of up to 17-inch increments with AQI increases of 2.48 percent; ADTFL increases of 1 percent with AQI increases of 0.07 percent; and an increase of 1 percent in FLTS or FLTDEL with AQI increases of 0.04 percent (assuming POP and MANUF increases of 1 percent). An increase in the number of flights in the U.S. by 25 percent equates to about a 1 percent increase in AQI which is less than a 1-unit increase.

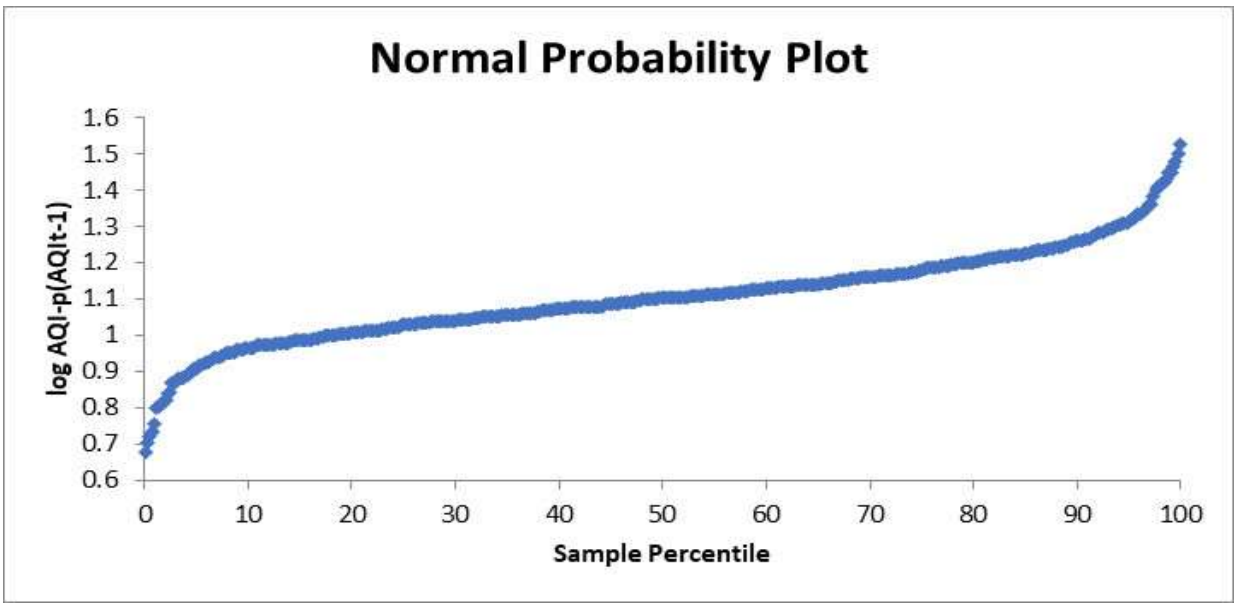
Figure 2 – Log of Air Quality Index Regression

```
. regress logaqipaqit1 mtns rain logadtflpadtflt1 logpopmanuffltsfltdelppopmanuffl

      Source |       SS           df       MS      Number of obs   =       468
-------------+----------------------------------   F(4, 463)       =     33.52
       Model |  1.6859544         4  .4214886      Prob > F        =    0.0000
    Residual |  5.82165042       463  .012573759   R-squared       =    0.2246
-------------+----------------------------------   Adj R-squared   =    0.2179
       Total |  7.50760482       467  .016076242   Root MSE        =    .11213

------------------------------------------------------------------------------
                 logaqipaqit1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                         mtns |   .0566673   .0114584     4.95   0.000     .0341503    .0791843
                         rain |  -.0247771   .0056837    -4.36   0.000    -.0359461   -.0136081
                logadtflpadtflt1 |   .0747645   .0460912     1.62   0.105    -.0158094    .1653383
logpopmanuffltsfltdelppopman~l |   .0438567   .0078137     5.61   0.000     .0285019    .0592115
                        _cons |    .414796   .1337735     3.10   0.002     .1519176    .6776744
------------------------------------------------------------------------------

. ovtest

Ramsey RESET test using powers of the fitted values of logaqipaqit1
       Ho:  model has no omitted variables
                 F(3, 460) =      4.08
                  Prob > F =      0.0070
```

However, the results of the Ramsey RESET are indicative that the model may be misspecified as $F(3, 460) = 4.08$ is above the F table threshold of about 2.64 and Prob > F = 0.0070, which is < 0.05. The use of quadratic IVs did not improve the model. It was found that corrections to several likely inaccurate data points in addition to removal of the RAIN and ADTFL IVs brought the Ramsey RESET output to the 2.64 and 0.05 thresholds. Under this model the overall F value improves to 54.68, the *F statistic* remains statistically significant, and the R^2 drops to 0.19. The MTNS and (POP+MANUF)*(FLTS+FLTDEL) coefficients increase to 0.075 and 0.050, respectively, while statistical significance of *t Stat* and *p-values* improves slightly. Nevertheless, the decision was made to keep the RAIN and ADTFL IVs, to retain diversity of AQI predictability with the understanding there could be some loss in accuracy. Thus, the remainder of the paper reflects this decision.

The four standard regression assumptions used for validation are: 1) normal distribution of variables without skewing or excessive outliers; 2) the IVs and DVs have a linear relationship; 3) variables have been measured dependably without error; and 4) homoscedasticity or consistency of the IV errors based upon scatterplot analysis. The normal probability plot in *Figure 3* below is relatively linear but the tailing at both ends indicates to an extent that the data is not normal. The second assumption can be addressed by analyzing the residual plots of the standardized residuals as a function of predicted values to see if non-linear patterns are prevalent (*Figure 4*). These residuals appear to be linear to an extent, although there are a few outliers that are more prevalent in the mid to left hand sections.

Figure 3



Nevertheless, the residuals are fairly well distributed and average very close to 0. *Figure 5* depicts a scatterplot of the residuals against the aircraft flight-related IV which shows relatively similar results. *Figure 6* shows a line fit plot of both the actual and the predicted Y_t^*against the aircraft-flight-related IV. The actual data display shows a slight coning effect opening to the left

while the predicted plots show a much more tightly and linear spread. *Figures 3-6* are indicative of a possible concern with heteroscedasticity.

Figure 4

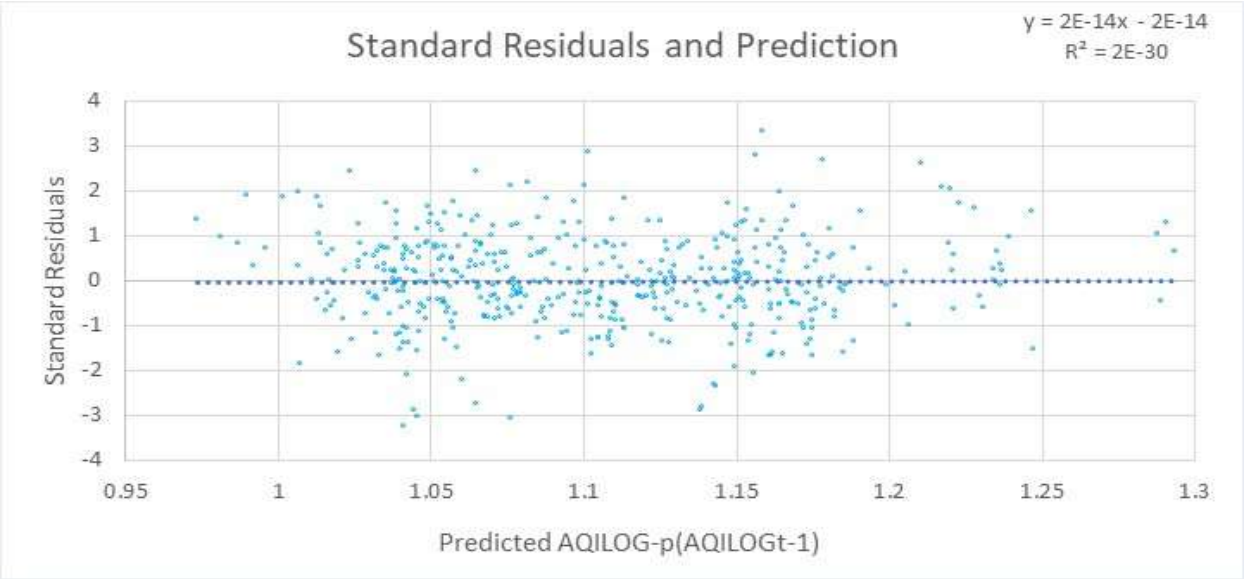


Figure 5

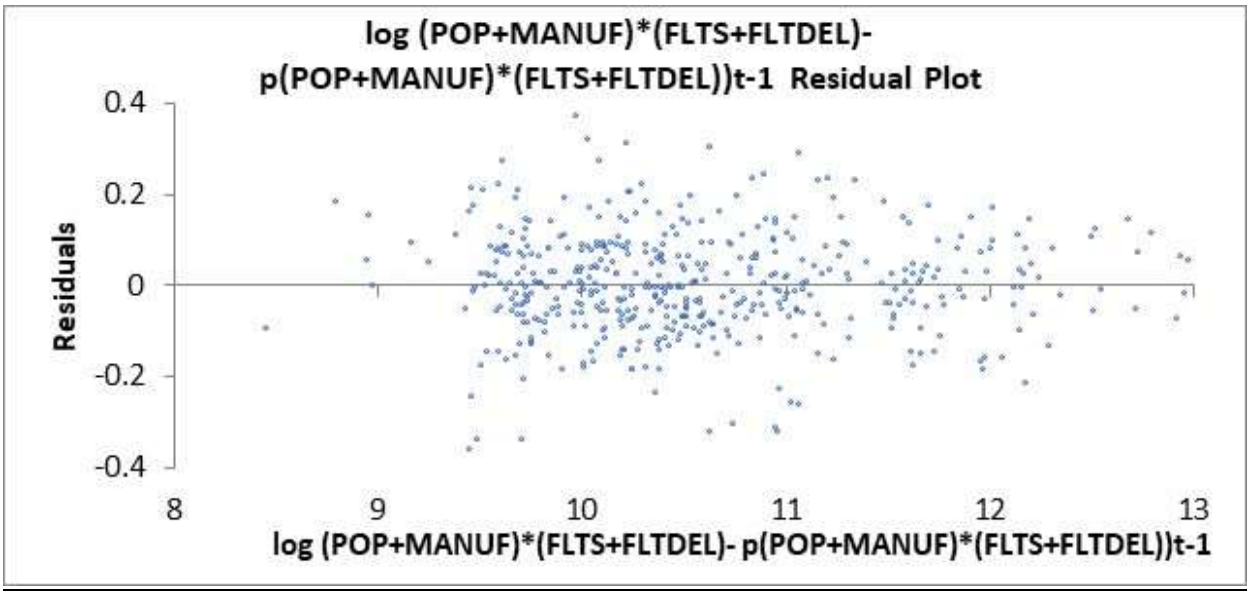
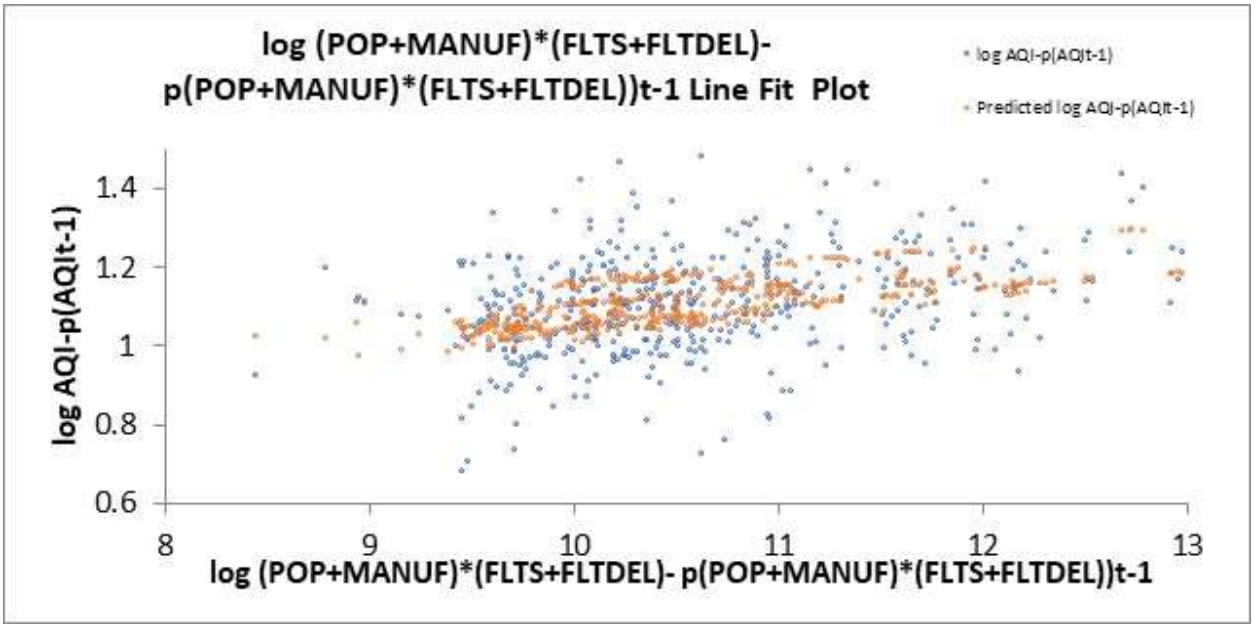


Figure 6



A Park Test was performed to check further for heteroscedasticity in the model by squaring the residuals or errors, converting them to log format, and regressing the predicted *Y* values against the results. The resulting R^2 is 0.0026, the *t stat* value is -1.0958 which is not above the t table threshold of 1.960, and the *P-value/Significance F* value is 0.2737 which is above 0.05. Consequently, there is not a statistically significant relationship and there is no evidence of heteroskedasticity in the model.

A possible concern is that the OLS estimated IV coefficients are determined pursuant to the mean values and only indicate average marginal effects on the DV. Therefore, quantile regressions were run at the 10th, 25th, 50th, 75th and 90th percentiles to identify different marginal effects of aircraft operations on AQI. The results are in *Figure 7* below and show that increases in aircraft flights and delays of 1 percent are fairly consistent with AQI increases ranging from 0.3 to 0.5 percent.

Figure 7: Estimated Quantile Regression Slope Coefficients

	OLS	Quantile Regression				
	Mean	10th Percentile	25th Percentile	Median	75th Percentile	90th Percentile
Intercept	0.4147959	0.2329844	0.3722686	0.4104510	0.6031181	0.6621140
	(0.1337735)	(0.2744162)	0.1619508	(0.1515747)	(0.1883509)	(0.2712881)
MTNS	0.0566670	0.5537170	0.4916050	0.0471406	0.0589491	0.0676838
	(0.0114584)	(0.0235053)	(0.1387200)	(0.0129832)	(0.0161333)	(0.0232373)
RAIN	-0.0247771	-0.0153349	-0.0246627	-0.2536010	-0.2287600	-0.0370497
	(0.0056837)	(0.0116592)	(0.0068809)	(0.0064400)	(0.0080025)	(0.0115263)
log ADTFL	0.0747645	0.0855759	0.0378340	0.0722296	0.0387851	0.0886739
	(0.0460912)	(0.0945491)	(0.0557996)	(0.0522225)	(0.0648956)	(0.0934713)
log (FLTS+DEL)	0.0438567	0.0419914	0.0542638	0.0453588	0.0446274	0.0325873
	(0.0078137)	(0.0160287)	(0.0094596)	(0.0088535)	(0.0110016)	(0.0158460)
Observations		468	468	468	468	468
Note: Dependent variable is log AQI-p(AQIt-1). Standard errors in parentheses. Highlighted values are significant at the 5 percent level.						

The effect of ADTFL increases of 1 percent on AQI range from 0.04 to 0.09 percent, however, none are statistically significant. Of note is that areas with MTNS can expect AQI increases ranging from 4.71 to 55.37 percent. Annual RAIN reductions of up to 17-inch increments are associated with AQI increases ranging from 1.53 to 25.36 percent.

Conclusions and Recommendations

The study found that a 1 percent increase in either number of flights, number of flight delays, or a combination of both is associated with an increase of the AQI within MSAs by 0.04 percent, assuming a comparable increase in population and manufacturing employment or a mixture of both. Thus, the elasticity is relatively small. The elasticity of ADTFL is 0.07, ceteris paribus. A recommendation is to identify RAIN and TEMP data by year in an attempt to improve the model. Wind speed and direction data would also likely be beneficial. Of note is that the study does not focus on those populations within very close proximity to airports that are likely more severely impacted by aircraft traffic. Therefore, a recommendation is to conduct more focused research on those data collection points with identification by distance to airports

to measure the incremental relationship with AQI. In a separate vein, results from the cumulative research can guide decisionmakers as to the appropriate prices that airline customers could be charged to account for the associated social costs of air pollution based upon known morbidity and mortality rates. The implementation of this and other aviation travel demand management (TDM) techniques may be appropriate to reduce air traffic. The extent of the price increases could be determined through a benefit-cost analysis study which would also capture the disproportionate externalities to those residing near high-capacity hub airports compared to other terminals. For example, in the year 2000, the Chicago metropolitan area had about 3 percent of the U.S. population, yet O'Hare Airport alone had almost 5 percent of the U.S. departures and with other Chicago airports the figure was 6 percent (Arkell, 2005).

References

Arkell, R. (2005). O'Hare Modernization Draft Environmental Impact Statement Comments. Retrieved via ftp://public-ftp.agl.faa.gov/OMP%20PFC%2006-19-C--00-ORD/EIS%20and%20ROD%20Administrative%20Record/Disk05/03_March%202005/18_02_02_14919.pdf.

Balakrishnan H, Hansman J. (2010). Reducing Airport Surface Operation Environmental Impacts. MIT Laboratory for Aviation and Environment. Retrieved via http://lae.mit.edu/reducing-airport-surface-operations-environmental-impacts/.

Barth M, Boriboonsomsin K. (2009). Traffic Congestion and Greenhouse Gases. Access. 35: Fall. Retrieved via http://www.accessmagazine.org/fall-2009/traffic-congestion-greenhouse-gases/.

Currie J, Walker R. (2011). Traffic Congestion and Infant Health: Evidence from E-Z Pass. American Economic Journal: Applied Economics 3. January: 65-90.

Gujarti D, Porter D, (2010). Essentials of Econometrics. 4th Edition. McGraw-Hill.

Hilmer C, Hilmer M. (2014). Practical Econometrics. McGraw-Hill.

Lin S. Munsie J. Herdt-Losavio M. Hwang S. Civerolo K. McGarry K. Gentile T. (2008). Residential proximity to large airports and potential health impacts in New York State. Int. Arch. Occup. Environ. Health. 81: 797-804.

Lu C. (2011). The economic benefits and environmental costs of airport operations: Taiwan Taoyuan International Airport. Journal of Air Transport Management. 17: 360-363.

Ratliff G, Sequeira C, Waitz I, Ohsfeldt M, Graham M, Thompson T, Thrasher T. (2009).

Aircraft Impacts on Local and Regional Air Quality in the United States. PARTNER Project 15 final report. Partnership for AiR Transportation Noise and Emissions Reduction.

Schlenker W, Walker W. (2012). Airports, Air Pollution, and Contemporaneous Health. University of California at Berkeley.

Shirmohammadi F, Sowlat M, Hasheminassab S, Saffari A, Ban-Weiss G, Sioutas C. (2017). Emission rates of particle number, mass and black carbon by the Los Angeles International Airport (LAX) and its impact on air quality in Los Angeles. Atmospheric Environment. 151:82-93.

Appendices

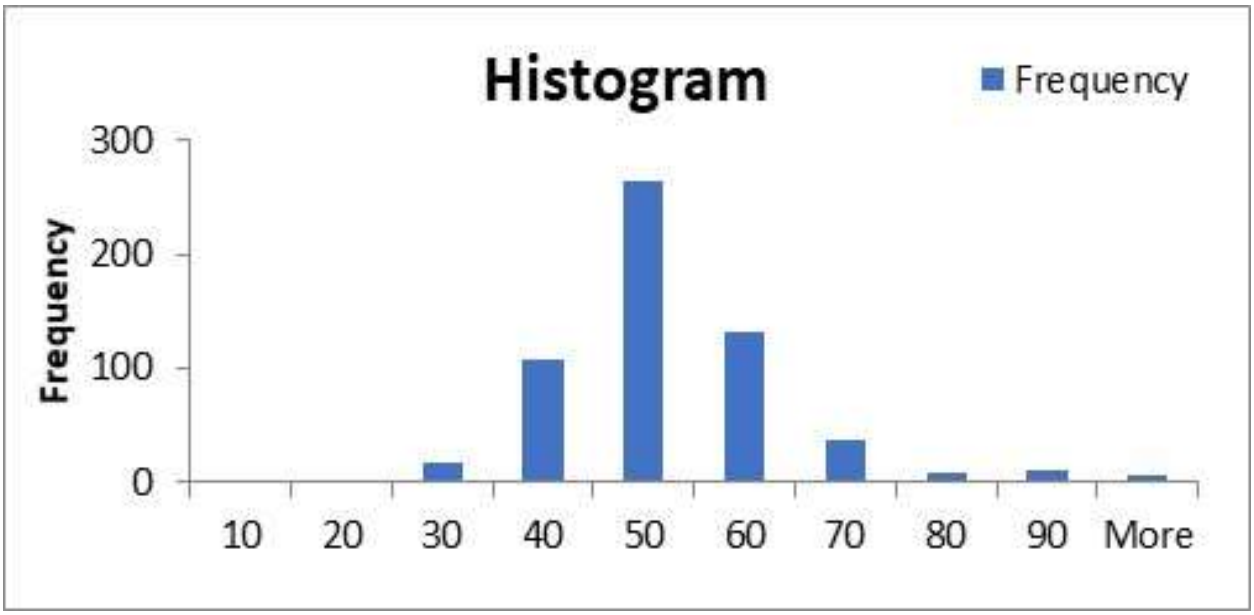
Appendix 1 – Pearson Correlation Matrix

	PLNS	MTNS	CPLL	NE	COAL	TEMP	RAIN	10WPD	WPDCHG	WPD	POP	GDP	PCGDP	MANUF	FLTS	PASS	FLTDEL	PASS PER FLT	FLTS+FLTDEL	POP*(FLTS +FLTDEL)	(POP+MANUF)* (FLTS+FLTDEL)	ADTFL	AQI
PLNS	1																						
MTNS	-0.32977	1																					
CPLL	-0.45175	-0.33637	1																				
NE	-0.18355	-0.13939	-0.06081	1																			
COAL	0.251051	-0.00886	-0.20953	-0.14971	1																		
TEMP	-0.4707	0.035247	0.53928	-0.23817	-0.21527	1																	
RAIN	-0.01563	-0.31158	0.202456	0.206956	0.185031	0.061894	1																
10WPD	-0.13942	-0.05717	0.266361	0.283331	-0.20941	0.059877	-0.17314	1															
WPDCHG	-0.2441	0.042708	0.097804	0.10049	-0.06084	0.053801	-0.01978	-0.11559	1														
WPD	-0.14722	-0.05605	0.270123	0.287227	-0.21189	0.061684	-0.17426	0.99955	-0.08575	1													
POP	-0.0533	-0.05028	0.210048	0.199725	-0.09826	0.081797	0.04124	0.80366	-0.20355	0.799958	1												
GDP	-0.05281	-0.08812	0.219965	0.236556	-0.10258	0.039417	0.050376	0.836733	-0.19664	0.83334	0.983299	1											
PCGDP	-0.18092	0.146409	0.108851	-0.17641	-0.11669	0.251527	-0.17505	-0.24239	0.071875	-0.24096	-0.29385	-0.35746	1										
MANUF	0.112951	-0.04068	0.068368	0.092009	-0.01562	-0.06096	0.025047	0.594515	-0.27409	0.588048	0.874023	0.85054	-0.36055	1									
FLTS	-0.0865	-0.02878	0.203205	0.074588	-0.12611	0.145527	-0.00058	0.67288	-0.20361	0.668778	0.89332	0.870008	-0.3266	0.794639	1								
PASS	-0.08234	-0.05895	0.206778	0.080928	-0.11743	0.123022	0.027638	0.647945	-0.20154	0.64383	0.860414	0.843322	-0.3501	0.754665	0.977712	1							
FLTDEL	-0.0074	-0.03683	0.096295	0.038427	-0.05278	0.066238	0.007129	0.500614	-0.28876	0.493418	0.740922	0.722284	-0.31652	0.700415	0.887996	0.915857	1						
PASS PER FLT	-0.03162	-0.06566	0.159456	0.044796	0.008223	0.097664	0.105523	0.38149	-0.21202	0.376249	0.541678	0.531934	-0.46882	0.521666	0.635347	0.713159	0.643511	1					
FLTS+FLTDEL	-0.08475	-0.02906	0.201083	0.073886	-0.12462	0.143939	-0.00038	0.670407	-0.2063	0.666216	0.891949	0.86869	-0.32724	0.794455	0.999934	0.978841	0.893237	0.63729304	1				
POP*(FLTS+FLTDEL)	-0.0492	-0.05916	0.19579	0.184171	-0.11294	0.031337	0.020641	0.828674	-0.19765	0.825226	0.93714	0.942188	-0.22013	0.781647	0.837575	0.791004	0.678193	0.35550147	0.8358765	1			
(POP+MANUF)*(FLTS+FLTDEL)	-0.04775	-0.05842	0.194966	0.182114	-0.11268	0.030722	0.019591	0.826484	-0.19945	0.822975	0.937828	0.942364	-0.2209	0.786547	0.839263	0.792178	0.68052	0.35635816	0.8375852	0.999946	1		
ADTFL	-0.2951	0.130412	0.311067	0.0779	-0.09924	0.268018	-0.00414	0.424607	-0.01849	0.425339	0.539524	0.50211	-0.29043	0.516983	0.516131	0.525164	0.455585	0.53139335	0.5160286	0.357287	0.359188781	1	
AQI	-0.09907	0.403779	-0.11053	-0.06066	0.002519	0.184863	-0.22438	0.231824	-0.15315	0.227905	0.410239	0.335622	-0.03934	0.412932	0.335717	0.299526	0.311107	0.22490327	0.3360205	0.270846	0.272824935	0.451319	1

Appendix 2 – Covariance Matrix

	PLNS	MTNS	CPLL	NE	COAL	TEMP	RAIN	10WPD	WPDCHG	WPD	POP	GDP	PCGDP	MANUF	FLTS	PASS	FLTDEL	PASS PER FLT	FLTS+FLTDEL	POP*(FLTS +FLTDEL)	(POP+MANUF)* (FLTS+FLTDEL)	ADTFL	AQI
PLNS	0.202498																						
MTNS	-0.07188	0.234641																					
CPLL	-0.09643	-0.07729	0.224998																				
NE	-0.02761	-0.02257	-0.00964	0.111769																			
COAL	0.103879	-0.00394	-0.09139	-0.04602	0.845496																		
TEMP	-0.20754	0.016729	0.250639	-0.07802	-0.19395	0.960041																	
RAIN	-0.00679	-0.14581	0.092775	0.066842	0.164366	0.058587	0.933304																
10WPD	-207.086	-91.4001	417.0311	312.6536	-635.582	193.6491	-552.104	10894762															
WPDCHG	-10.913	2.055293	4.609007	3.337673	-5.55759	5.237212	-1.89816	-37904.1	9870.208														
WPD	-217.999	-89.3448	421.6402	315.9912	-641.14	198.8863	-554.002	10856858	-28033.9	10828824													
POP	-62067.6	-63016.7	257807.7	172774.5	-233791	207381.9	103091.4	6.86E+09	-5.2E+07	6.81E+09	6.7E+12												
GDP	-4364.28	-7838.31	19160.59	14523.14	-17322.2	7092.498	8937.221	5.07E+08	-3587588	5.04E+08	4.67E+11	3.37E+10											
PCGDP	-0.50943	0.443779	0.323087	-0.36905	-0.67143	1.542145	-1.05822	-5006.4	44.68265	-4961.71	-4757801	-410758	39.15552										
MANUF	3814.267	-1478.73	2433.627	2308.335	-1077.93	-4482.61	1815.835	1.47E+08	-2043469	1.45E+08	1.7E+11	1.17E+10	-169307	5.63E+09									
FLTS	-10200.4	-3653.33	25260.37	6534.961	-30390.2	37368.37	-145.582	5.82E+08	-5301212	5.77E+08	6.06E+11	4.19E+10	-535585	1.56E+10	6.87E+10								
PASS	-401788	-309636	1063622	293393.6	-1170939	1307143	289547	2.32E+10	-2.2E+08	2.3E+10	2.41E+13	1.68E+12	-2.4E+07	6.14E+11	2.78E+12	1.18E+14							
FLTDEL	-22.3776	-119.824	306.8241	86.29643	-325.996	435.9643	46.26313	11099655	-192705	10906951	1.29E+10	8.91E+08	-13304.5	3.53E+08	1.56E+09	6.67E+10	45122801						
PASS PER FLT	-0.19959	-0.44608	1.060879	0.210053	0.106056	1.342191	1.429862	17661.47	-295.448	17366.02	19659077	1370118	-41.1469	549081.2	2335403	1.08E+08	60630.14	196.72923					
FLTS+FLTDEL	-10222.8	-3773.15	25567.2	6621.258	-30716.2	37804.34	-99.3188	5.93E+08	-5493917	5.88E+08	6.19E+11	4.28E+10	-548890	1.6E+10	7.02E+10	2.85E+12	1.61E+09	2396032.94	7.1852E+10				
POP*(FLTS+FLTDEL)	-7.7E+10	-1E+11	3.23E+11	2.14E+11	-3.6E+11	1.07E+11	6.93E+10	9.5E+15	-6.8E+13	9.43E+15	8.42E+18	6.01E+17	-4.8E+12	2.04E+17	7.63E+17	2.98E+19	1.58E+16	1.7321E+13	7.7833E+17	1.207E+25			
(POP+MANUF)*(FLTS+FLTDEL)	-7.6E+10	-1E+11	3.29E+11	2.17E+11	-3.7E+11	1.07E+11	6.73E+10	9.7E+15	-7E+13	9.63E+15	8.63E+18	6.16E+17	-4.9E+12	2.1E+17	7.82E+17	3.06E+19	1.63E+16	1.7782E+13	7.9872E+17	1.236E+25	1.2656E+25		
ADTFL	-459.821	218.7379	510.9155	90.17813	-315.982	909.3128	-13.8454	4852893	-6361.16	4846532	4.83E+09	3.19E+08	-6292.69	1.34E+08	4.68E+08	1.97E+10	10596758	25808.0779	478958634	4.298E+15	4.42463E+15	11989752	
AQI	-0.5087	2.231704	-0.59822	-0.23138	0.02643	2.06674	-2.4734	8730.88	-173.611	8557.269	12111992	703245.5	-2.80854	353573.2	1003878	37061188	23845.11	35.9932564	1027722.62	1.074E+13	1.10745E+13	17831.15	130.1913

Appendix 3 – Air Quality Index Data Histogram



Appendix 4a – Stata Negative Binomial Model Run[1]

```
. mfx

Marginal effects after nbreg
      y  = Predicted number of events (predict)
         =  13.084001
```

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
mtns*	1.98136	.38814	5.10	0.000	1.22061	2.74211		.376068
rain	-.77973	.18477	-4.22	0.000	-1.14188	-.417579		2.47863
adtflp~1	.0006777	.00015	4.43	0.000	.000378	.000977		3833.58
popman~f	5.58e-07	.00000	3.73	0.000	2.6e-07	8.5e-07		302296

(*) dy/dx is for discrete change of dummy variable from 0 to 1

[1] Note that the last IV data was divided by 1 million due to this Stata error: warning: derivative missing; try rescaling variable popmanuffltsfltdelppopmanuffltsf.

Appendix 4b – Stata Negative Binomial Model Run

```
. *(5 variables, 468 observations pasted into data editor)

. nbreg aqipaqit1 mtns rain adtflpadtflt1 popmanuffltsfltdelppopmanuffltsf
note: you are responsible for interpretation of non-count dep. variable

Fitting Poisson model:

Iteration 0:   log likelihood = -1263.9809
Iteration 1:   log likelihood = -1263.9805
Iteration 2:   log likelihood = -1263.9805

Fitting constant-only model:

Iteration 0:   log likelihood =  -1693.284
Iteration 1:   log likelihood = -1317.2842
Iteration 2:   log likelihood = -1316.9983
Iteration 3:   log likelihood = -1316.9548
Iteration 4:   log likelihood = -1316.9547

Fitting full model:

Iteration 0:   log likelihood = -1269.6338
Iteration 1:   log likelihood = -1264.2588
Iteration 2:   log likelihood = -1263.6611
Iteration 3:   log likelihood = -1263.6248
Iteration 4:   log likelihood = -1263.6241
Iteration 5:   log likelihood = -1263.6241

Negative binomial regression                      Number of obs     =        468
                                                  LR chi2(4)        =     106.66
Dispersion     = mean                             Prob > chi2       =     0.0000
Log likelihood = -1263.6241                       Pseudo R2         =     0.0405

------------------------------------------------------------------------------
                        aqipaqit1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------------------------+-------------------------------------------------------------
                             mtns |   .1485351   .0285752     5.20   0.000     .0925287    .2045415
                             rain |  -.0595942   .0141384    -4.22   0.000    -.0873048   -.0318835
                      adtflpadtflt1 |   .0000518   .0000117     4.43   0.000     .0000289    .0000747
popmanuffltsfltdelppopmanuffltsf |   4.26e-08   1.14e-08     3.73   0.000     2.02e-08    6.50e-08
                            _cons |   2.451801     .05923    41.39   0.000     2.335713     2.56789
----------------------------------+-------------------------------------------------------------
                         /lnalpha |  -5.543513   1.232474                     -7.959118   -3.127908
----------------------------------+-------------------------------------------------------------
                            alpha |   .0039128   .0048224                      .0003495    .0438093
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 0.71                 Prob >= chibar2 = 0.199
```