IPAT- Fuzzy Model in Measuring Air Pollution: Evidence from Malaysia

Mohamad Idham Md Razak
Economic Department
Faculty of Business Management
Universiti Teknologi MARA
Alor Gajah Campus

Ismail Ahmad
Professor,PhD,Finance Department
Faculty of Business Management
Universiti Teknologi MARA
Shah Alam Campus

Imbarine Bujang
PhD,Economic Department,
Faculty of Business Management
Universiti Teknologi MARA
Sabah Campus

Adi Hakim Talib,
Statistics Department
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
Alor Gajah Campus

Zakimi Ibrahim
Faculty of Business Management
University of Technology MARA (UiTM)
78000 Alor Gajah, Melaka, Malaysia

Abstract

Air pollutants can be either gases or aerosols which particles or liquid droplets suspended in the air. They change the natural composition of the atmosphere, can be harmful to humans and other living species and can cause damage to natural water bodies and the land. Anthropogenic specifically due to the human causes that in this study, it has been identified that Population, Gross Domestic Product (GDP) and Manufacturing Industry adaptive from IPAT Model is the major contributors to the emission of carbon dioxide. The time series data gained and analyzed from the year 1970 to 2011 to explain the relationship among the variables. From the time series analysis, the results are statistically significant and improved after transforming into fuzzy numbers and free form autocorrelation, multicollinearity and heterokedasticity problem. Hence, the information provided will assist the government in the future planning and development.

Keyword: Air Pollution, IPAT Model, distance-Based Ranking Fuzzy Numbers Approach

Introduction

A change in economic structure, which is seen as the growth of the manufacturing industry, affects the amount carbon emissions. This happens as a particular economy shifts from subsistence to an agricultural economy and later, to an industrial economy.

In this case pollution levels increase as the result the significant as well as instrumental changes in the manner of the production processes. Moreover, the trade policy which comes under the major activity in Gross Domestic Product (GDP) stipulates that the more open an economy, the greater the possibilities for importing and exporting pollution intensive commodities, and to another end, the lower or higher the domestic pollution levels.

Another matter that warrants attentions is the condition of economic scale. It considers the magnitude of a population; the bigger the economy, the greater the pollution, and this happens as everything else remains constant. These three questions describe the problem statement of this following research. This research investigates the impacts of population density, Gross Domestic Product (GDP) and the manufacturing industry with regards to the links between the growth of population and the environment. Furthermore, such investigation also contributes to the argument that the size of the population and other determinants must be given due consideration in the forecasting of air pollutant emissions.

This study is therefore, very much related to the vast and dynamic literature related to the Environmental Kuznets Curve (EKC). The EKC states that pollution with regards to the environment increases and decreases along the rise in per capita income levels. Examples of this can found in the research done [4], [10] all of whom analyzed the impacts highlighted. Population is often included only as a scale variable in EKC studies. There are very few systematic quantitative empirical studies on the relationship between population and pollution that are explicitly examined. [6] studied the effects of population magnitude on the level of air pollution in the U.S state of California came up with conclusion that some sources of emissions are closely related to the population while others not.

However, the global implications of Cramer's work is not adequate; this is due to them limiting their attention on only one state in a developed country as the U.S and as such, their main result is far from robust. [8] together with [18] in their studies which focus on the emissions of CO2 and energy use, investigated the roles played by population, affluence and technology by adopting the Impact-Population-Affluence-Technology (IPAT) model. They found out that the ratio of change between energy use and CO2 emissions is bordering on unity. For example, it was estimated that an approximately one percent increase in CO2 emissions is the result of one percent increase in population. However, [7], did not make projections as to how such elasticity might differ with population levels. Moreover, these results are derived from a one year- cross-sectional study. [7], utilizing IPAT model too, employs a group of cross-sectional and time series data for similar investigation. They found the population elasticity for CO2 to be around 1.41 and 1.65.

But, he also did not study how different population size might affect the kind of result that is to be obtained from such study. While being on the right track, [7] research has potentially severe methodological and justification flaws as his study concerns only one pollutant, which is CO2. For instance, their variables, namely the CO2 emissions and per capita income, show consistent increase from time to time. Thus, the covariance stationary condition required for non-biased and regression-stable outcomes are nonexistent. Hence, the validity of the estimated coefficients and elasticity is questionable. Moreover [11], noted that many economists have undertaken studies on ‘‘Environmental Kuznets Curve’’ (EKC) by including population density as one of many of determinants of pollution concentrations.

Yet, those studies often to produce inconsistent results. The investigation on the population–pollution relationship is also extensively dealt with in those studies, as well as the study on the wider effects of population levels against various other pollution related demographic factors or spatial density. This research try to provide an in depth elaboration on the effects of population magnitude and factors such as Gross Domestic Product (GDP) and manufacturing industry on air pollution emissions as well as to adjust the flaws of previous studies outlined above. This research is rooted on the works by [7] and [18] and seeks to better their studies in many ways. First, the three studies mentioned earlier investigate only CO2 and energy use. This particular research on the other hand, consolidates the discussion by forecast the value and comparing the consistency and findings of the irregularities that occur along the period of study.

Secondly, unlike [7] and [18] the researcher presents a time-series data analysis. The justification is that it is able to record changes from time to time. It also allows for more elaborate research design that grants effective controlling of the ‘latent country’ effects.Not only that, comparison of the results with the adaption of ranking fuzzy approach can be achieved.

Third, the researcher, using a first-differenced estimator, rectifies the methodological flaw of [7] investigation by keeping his variables are co-variance stationary. The researchers estimated results are therefore consistent and not biased.

Theoretical background and literature reviews

The scope of the literature is limited to three variables based on the IPAT Model proposed by [5] and [9]. Most of the researchers claimed their models to be the best and come with minimum error. The IPAT model is based on the equation I = f (P, A, T), in which I represents environmental impact with *P*, *A* and *T* represented by the variables population, affluence and technology respectively. Global population keeps increasing year by year. There are numerous literature found that there is a direct correlation between increasing number of population that lead to high air pollution emission.[13], came up with the empirical evidence an increase in carbon dioxide emissions is positively associated with global population change. The result was derived from the analyzing of the links between carbon dioxide emissions, population, and other related determinants.

His based his empirical study is on a cross-country data involving 83 countries from the year 1980 to 2007. Results obtained validated the notion that population dynamics affects the amount of carbon dioxide emissions. [13] also added that the results has confirmed that population is an instrumental factor in the manipulation of carbon dioxide increase. In Malaysia, tools for transportation as motorcycles, cars, vans, omnibuses, trucks as well as heavy machineries such as tractors operate on fuel. Usually, the heavy ones operate on diesel. Light vehicles on the other hand, operate on petrol. [12] posits that diesel consuming vehicles contributes more to air pollution. This is because more road dirt, NOx and SOx and other particles are produced when compared to vehicles consuming petrol.

The Department of Education measured that about 622,000 tonnes of air pollutants are released into the atmosphere. These pollutants, by large are by products of vehicle fuel combustion [15]. 48.7 percent of the pollutants are carbon monoxide particles (CO), with the rest being SO2 (31.3 percent), NO2 (11.2 percent), hydrocarbon (6.0 percent) and unstable particles (2.8 percent) [15]. About 96 percent of hydrocarbons come from cars, motorcycles, aircrafts and rail transport [15]. He also noted that transportation industry is also estimated to contribute to 70 percent of the total NOx generation by fuel combustion in Malaysia. According to the research conducted by [14], he was numerous efforts to derive the EKC theoretically. The main theoretical explication is that when the number of GDP grew, the bigger scale of production and this in turn, causes to more pollution to take place.

However, high income per capita, and the subsequent demand for better health and environmental wellbeing can translate into environmental checks, in this case there will be agents for favorable shifts in the composition of output and in production techniques. Malaysia, known to be a developing country throughout the world, enjoys a healthy economy growth, as can be seen the thriving of many economic activities. The relationship between Malaysia's GDP and her Carbon Emission Index can only be justified by performing certain test that will be explained in the next chapter. The paper by [17] has already provided a prelude into this matter. [17] chose Malaysia as the subject of his study because Malaysia has achieved a lot of economic progress, apart from of it being one of the fastest growing economies in the region as well as in the world for decades.

Not only that, according to him also Malaysia is a good to study because it has more and quality data on environmental quality than perhaps any other country. [17] considered his study to being groundbreaking on the matter of the pollution/income relationship over time for a developing country. In Malaysia too, a big portion of solid wastes are disposed or discarded using the landfill system. The wastes disposal sites, which often look like huge reactors, are normally filled up with wastes. The composting processes of the organic materials that occur as the microorganisms in the soil started to break down the wastes. The process produces several gases including methane (CH4), CO2, CO, H2S and vinyl chloride as a result, which further worsen the air condition. Methane is the gas that is generated in the biggest amount. It is highly flammable [2] and in some instances may cause explosion and this have somehow indirectly play major role in the air pollution mechanism.

Methodology and Sources of Data

Te empirical model or model regression that involve in this study consists three versions which are, original equation adopted from IPAT model, after the log and increase the robustness after fuzzy the model.

General equation:

$Y_{jt} = B_0 + B_1x_{1t} + B_2x_{2t} + B_3x3_t + e_t$ (1)

Adaptive equation:

$C0_2 = B_0 + B_1Pop_t + B_2GDP_t + B_3Man_t + e_t$ (2)

After log:

$\text{Log } C0_2 = B_0 + \text{Log } B_1Pop_t + \text{Log Log } B_3Man_t + e_t)$ (3)

After Fuzzy:

$\text{Log Fuzzy } C0_2 = B_0 + \text{Log Fuzzy } B_1Pop_t + \text{Log Fuzzy } B_2GDP_t$
$+ \text{Log Fuzzy } B_3Man_t + e_t)$ (4)

In which, y_{jt} represent the Air Pollution trend represented by Carbon Emission that involve in this study which are Pollution, Gross Domestic Product and Manufacturing Industry. In the time series analysis, this study will conducting several important tests and choose the best model estimator to explore the impact of Pollution-Afflunce-Technology (PAT) variables structures on carbon emission. Test for stationary (unit root), multicollinearity, heteroscedasticity and autocorrelation will be conducted using the recent and sophisticated techniques or methods to know the problem of time series data. To be more emphasizing, ranking fuzzy numbers has been used widely in various decision-making problems that involve subjectivity or impreciseness.

Essentially, all the presented ranking methods try to discriminate fuzzy numbers correctly especially for the cases of complex overlapping fuzzy numbers and also solving some combinations where previous methods cannot rank correctly. Later, many ranking fuzzy numbers methods have been proposed. Some of the approaches among others are based on area compensation, induced function using weighted average of fuzzy numbers, ranking using distance minimization, α- cuts, belief feature and signal/noise ratio, lexicographic screening procedure, ranking fuzzy numbers using radius of gyration, ranking fuzzy numbers based on heights and spreads, ranking method using deviation degree and centroid-based technique [1].

$$\Delta(\tilde{A}) = x_{\tilde{A}}^{*} \cdot h_{\tilde{A}}^{*} \cdot \varphi^{*}(\hat{\delta}_{\tilde{A}\tilde{B}}, \tilde{A})^{\sigma^{*}(\tilde{A})} \quad (5)$$

Empirical Result

The usual test statistic (t test and F test) would not have the standard distribution, when a model is consisted of non- stationary variables, Thus, before proceeding to estimating the model, it is important that non-stationary tests on variables should be carried out. A time series that is non-stationary can be converted to a stationary series if it can be distinguished appropriately.

Table 1: Test for Stationary Data (Actual Value)

Variables	Level	Stationary / Non-stationary	First Difference	Stationary / Non-stationary
POL	0.6989	Non-stationary	0.0600	Non-stationary
POP	0.5544	Non-stationary	0.0560	Non-stationary
GDP	0.1131	Non-stationary	0.1000	Non-stationary
MAN	0.9208	Non-stationary	0.6720	Non-stationary

Table 2: Test for Stationary Data (Fuzzy Value)

Variables	Level	Stationary / Non-stationary	First Difference	Stationary / Non-stationary
POL	0.6989	Non-stationary	0.0000	Stationary
POP	0.5544	Non-stationary	0.0000	Stationary
GDP	0.1131	Non-stationary	0.0300	Stationary
MAN	0.9208	Non-stationary	0.0000	Stationary

This analysis is to identify the different level between independent and dependent variables. At the level all the data can be regarded as non-stationary because more than significant value which is 0.05. Table 1 using the actual value shows that the data is not stationary at the first different and the data is stationary at the first difference as stated in Table 2 using fuzzy value. [18] exclusive study of CO2, warrant serious severe methodological and justification flaws. The various variables used by [18], in particular the per capita income and CO2 release, show consistent positive increase over time. As such, they are no covariance stationary condition needed for non-biased and consistent regression results thereby raising question marks over the validity of the measured coefficients and elasticity.

Since the data should be stationary, the actual data will be transformed into fuzzy value since there is the condition of uncertainty occurs in analyzing air pollution relationship in line with the study from [1] found that uncertainty will lead to bias results. Thus, the after transforming actual value into fuzzy numbers and gathered the stationary data at the first different, then the data able to proceed using a range of regressions to test for a significant relationship among the variables.

Table 3: Residual Test (Actual Value)

	Minimum	Maximum	Mean	Standard Deviation	N
Predicted value	29962.318	80784.914	51807.156	17034.70460	42

Table 4: Residual Test (Fuzzy Value)

	Minimum	Maximum	Mean	Standard Deviation	N
Predicted value	29962.318	80784.914	51807.156	17034.70460	42

Table 3 and Table 4 are the result of Residual tests which one using the actual value and the one using fuzzy value. Since the results are the same then the interpretation will be the same. With the data series of pollution rate covered for 42 years data from1970 to 2011, it can be concluded the residual statistics for the pollution rate. The pollution rate is between 2.99 percent and 8.07 percent for the 42 years respectively. Therefore, it means that the lowest pollution rate is at 2.99 percent and the highest pollution rate in Malaysia is at 8.07 percent for those 42 years. In addition, the mean of the pollution rate is at 5.18 percent and the standard deviation is at 1.70 percent.

Moreover, it also means that the average pollution rate can be predicted in Malaysia is only 5.18 percent more or less with 1.70 percent. Autocorrelation, which is also called as serial correlation, occurs when the error term observation and regressions are correlated. The theoretical error term **e** is a random variable that is part of the regression model, even before it is estimated. This error term represent a random 'shock' to the model, or something that is missing from the model. However we can never see the actual error term **e**. Therefore we use the pattern; this pattern is evidence of autocorrelation.

Table 5: Durbin Watson (Actual Value)

Durbin Watson
0.896634

Table 6: Durbin Watson (Fuzzy Value)

Durbin Watson
2.346873

To test whether the auto correlation problem exists or not in this equation, it will involve the Durbin Watson analysis. If the DW test is close to 0 it means that the estimated regression equation has a negative autocorrelation problem. However, if the DW test is close to 4 it means that the estimated regression equation has a positive autocorrelation problem. Nevertheless, if the DW test is close to 2 so it means that the estimated regression equation is free from autocorrelation problem. Table 5 shows with the actual value; there is an existence of positive serial correlation and to enhance the result then the actual data will be transformed into fuzzy numbers and its shows there is no serial autocorrelation problem as stated in Table 6. The Variance Inflation Factor (VIF) measures multi-co linearity by regression one independent variable on all of the remaining independent variables. To use the VIF to look for any possible multicollinearity, we run the regression, one for each independent. The result is as follows:

Table 7: Variance Inflation Test (Actual Value)

Variable	VIF
Population	6.659
GDP	7.083
Manufacturing Industry	9.600

Table 8: Variance Inflation Test (Fuzzy Value)

Variable	VIF
Population	3.659
GDP	1.083
Manufacturing Industry	3.600

The regression equation considers there is an existence of multicollinearity issue occurs here as shown in Table 7 if it is based on the actual value. Thus, to increase the robustness the actual data will be transformed into fuzzy numbers and based on the Variance Inflation Test for Table 8, there is no existence of multicollinearity. Generally, if the error terms do not have constant variance, they are said to be heteroscedastic. On the other hand, errors may increase as the value of an IV increases. In this particular study, consider a model in which the total number of population per year is the IV and the carbon emission that caused air pollution is the DV. Population with low volume will emits relatively little on carbon emission, and the variations in population across such carbon emission will be small and vice, resulting in heteroscedasticity.

Note that, in this example, a high number of populations are a necessary but not sufficient condition for large emission of carbon. In addition, at any time a high value for an IV is a necessary but not sufficient condition for an observation to have a high value on a DV that is heteroscedasticity is likely.

Table 9: Chi Square (Actual Value)

Chi Square p-value
0.9976

Table 10: Chi Square (Fuzzy Value)

Chi Square p-value
0.0037

Moreover, measurement error can cause heteroscedasticity. Some data might provide more accurate result than others. Note that this problem arises from the violation of another assumption, that variables are measured without error. Breusch-Pagan and White Test tests the null hypothesis that the error variances are all equal versus the alternative that the error variances are a multiplicative function of one or more variables. A large chi-square would indicate that heteroscedasticity was present in Table 9. In this example, the chi-square value was small for Table 10, indicating heteroskedasticity was probably not a problem or at least that if it was a problem; it was not a multiplicative function of the predicted values.

Table 11:Significant Value (Actual)

Variable	T statistics	Probability
Constant	-2.597916	0.0133
Population	-0.349942	0.7283
GDP	10.6708	0.0000
Manufacturing	5.981571	0.0000

Table 12: Significant Value (Fuzzy)

Variable	T statistics	Probability
Constant	0.789854	0.4384
Log (Population)	-6.201304	0.0000
Log (GDP)	14.69138	0.0000
Log (Manufacturing)	2.780782	0.0112

Table 13: R-Squared

	Actual Value	Fuzzy Value
R-Squared	0.960654	0.957547

By comparing the result from Table 11, Table 12, and Table 13, it is obviously that the result after being fuzzy is more reliable and significant. Most statistical tests rely upon certain assumptions about the variables used in the analysis. When these assumptions are not met the results may not be deemed trustworthy, resulting in a Type I or Type II error, or over- or under-estimation of significance or effect sizes.As [1] notes, "Knowledge and understanding of the situations when violations of assumptions lead to serious biases, and when they are of little consequence, are essential to meaningful data analysis". But, observed by [1] not many articles reports, having tested projection of the statistical assesment, rely on them for drawing their conclusions.

Conclusion

The results of the regression analysis will undermine the true relationship between the variables If the connection between independent variables (IV) and the dependent variable (DV) is not linear. This undermining carries two risks. The two risks are increased chance of a Type II error for that IV, and in the case of multiple regression, an increased risk of Type I errors which is over-estimation for other IVs that share variance with that IV. Researchers such as [1] recommends that the best way to solve the problem is by enhancing the analysis by quantifying the uncertainty which here fuzzy situation, given the uncertainty that may lead to biased estimation. Given that it is possible that the threat of irreversible events creates more pollution, the following research show those reversible events, in overall situation, induce more conservation, the lesser pollution cases occur. The "general condition" is that both the dangerous rate of the occurrence date and the penalty the event inflicts are non-decreasing functions of the pollution level.

Formally, this matter can be viewed as events that are reversible, with a penalty that will just equal the value forgone due to occurrence. The function of the value, however, generally decreases with pollution whereas for reversible events a non-decreasing penalty function has been postulated. For the latter type of events, whereas for irreversible events pollution decreases the "penalty" and the conflicting trends may lead to less conservation, pollution increases both the danger rate and the penalty and both effects lead to more conservation. In this respect, the exogenous events considered here vary greatly according to the endogenous events, for which event uncertainty always entails more conservation.

References

[1]Ahmad Syafadhli Abu Bakar, Daud Mohamad, Nor Hashimah Sulaiman. (2012). Distance-Based Ranking Fuzzy Numbers. Advances in Computational Mathematics and its Applications (ACMA)146 Vo. 1, No. 3, World Science Publisher, United States

[2]Canter, L. W. (1985). Environmental Impact Assessment. University of Oklahoma, McGraw Hill Book Company.

[3]Cohen JE. (1995). How Many People Can The Earth Support? *New York: Norton*.

[4]Cole, M. A., A. J. Rayner and J. M. Bates. (1997). "The Environmental Kuznets Curve: An Empirical Analysis," *Environment and Development Economics*, 2, 401-416.

[5]Commoner, B., Corr, M., & Stamler, P.J. (1971). The causes of pollution. Environment. 13(3), 2-19.

[6]Cramer, J. C. (1998). Population growth and local air pollution: methods, models, and results. In W. Lutz, A. Prskawetz & W. C. Sanderson (Eds.), *Population and environment* (A supllement to Vol. 28). *Population and development review*, pp. 22-52. New York: population Council.

[7]Dietz, T. and Rosa , E.A. (1994). *Rethinking the environmental impacts of Population, Affluence and Technology.* Human Ecology Review. Summer/Autumn, 1.

[8]Dietz, T., & Rosa, E. A. (1997). Effects of population and affluence on CO_2 emissions. *Proceedings of the National Academy of Sciences*, 94(1), 175-179.

[9]Ehlirch, P.R., & Holdren, J.P. (1971). Impact of population and growth. *Science*, 171, 1212-1217.

[10]Grossman, G. M. and A. B. Krueger. (1991). "Environmental Impacts of a North American Free Trade Agreement," NBER Working Paper No.3914, *Cambridge: National Bureau of Economic Research*.

[11]Grossman, G. M. and A. B. Krueger (1994), "Economic Growth and the Environment," *Quarterly Journal of Economics*, 110, 353-377.

[12]Norhalim Bin Mohd Horany (1998). Kajian Pencemaran Udara: Kesan Jerebu Terhadap Masalah Kesihatan di Johor Bahru. *Universiti Teknologi Malaysia: PSM thesis*.

[13]Sanglimsuwan, K. (2012). The impact of population pressure on carbon dioxide emission: evidence from a panel-econometric analysis. *International Research Journal of Finance and Economic*s, 8 (2012), 90-93.

[14]Seldon, T. and D. Song. (1994). "Environmental Quality and Development: Is There a Kuznets Curve for Air Pollution Emissions?" *Journal of Environmental Economics and Management*, 27, 147-162.

[15]Sham Sani (1987). Urbanization and the Atmospheric Environment in the Low Tropics – Experiences From The Kelang Valley Region Malaysia. *Universiti Kebangsaan Malaysia, Bangi*.

[16]Stokey, N. (1998). "Are There Limits to Growth," *International Economic Review*, 39(1), 1-31.

[17]Vincent, J. R. 1997. Testing for environmental Kuznets curves within a developing country. *Environment and Development Economics* 2: 417–31.

[18]York, R., Rosa, E.A., & Dietz, T. (2003b). Footprints on the earth: the environmental consequences of modernity. *American Sociology Review*, 68(2), 279-300.