Economic Growth and Environmental Quality: A Non-Parametric Kernel Estimation of the Environmental Kuznets Curve[*]

Savas Alpay and Syed Mahmud

Department of Economics, Bilkent University, Bilkent, Ankara, Turkey

Abstract

The proposed inverted U-type relationship between environmental degradation and per capita income under EKC hypothesis has been re-examined in this paper. A non-parametric kernel estimation technique has been employed to obtain varying point estimates of partial derivatives of sulfur dioxide emissions with respect to per capita income. This technique does not impose any a priori functional restrictions in the empirical testing of EKC hypothesis, and avoids the modelling criticism raised in the literature. More importantly, we present a test of EKC hypothesis by decomposing data into residential, industrial and commercial areas of a city (which are further divided into centre-of-the-city and suburban areas). Our results suggest a qualified support for the validity of EKC hypothesis: it is statistically supported for observations corresponding to centre-of-the-city (either residential or industrial or commercial). The hypothesis does not seem to hold for the suburban areas. Finally, a distinction between coastal and off-coastal cities turned out to be important as indicated by the climatic knowledge.

---

[*] We are thankful to Gene Grossman and James Stock for their comments on this paper.

1. Introduction

Understanding the impact of economic development on the environmental quality is becoming increasingly important as general environmental concerns are making their way into main public policy agenda. The relationship between environmental quality and economic development has been empirically modelled through emissions-income relationship by many authors, and the outcome of most of these studies has been formulated by the so called Environmental Kuznets Curve hypothesis. The environmental Kuznets curve (EKC) hypothesis proposes that there is an inverted U-shape relation between environmental degradation and income per capita, or a U-type relationship between environmental quality and income per capita. This has been taken to imply that economic development will eventually undo the environmental impacts of the early stages of economic development.

The literature on this issue has developed rapidly over the last decade starting with the work by Grossman and Krueger [1991]. Grossman and Krueger [1995] extended their earlier study and published the most comprehensive work on EKC. Their analysis has included fourteen different indicators of environmental degradation including sulfur dioxide, smoke, heavy particles, the state of the oxygen regime in river basins, fecal contamination of river basins, and contamination of river basins by heavy metals such as lead, cadmium and arsenic. They have shown an inverted U-type relationship between per capita income and emissions of SO_2 and suspended particulates.

EKC hypothesis has also been tested by many others for different indicators of environmental degradation, such as deforestation, carbon emissions and municipal waste. Sulfur dioxide was among the most commonly used environmental degradation indicators and EKC hypothesis has been shown to hold mostly for sulfur dioxide emissions in the literature. A very recent survey of EKC studies on sulfur by Stern et al. [1998] lists several reasons why choice of sulfur is of interest. For example, for sulfur, simultaneity problem in the econometric analysis will be less important as compared to other indicators such as carbon dioxide and deforestation; substitution possibilities are larger for sulfur. We will focus on sulfur dioxide emissions in this study.

In addition to studies by Grossman and Krueger, one can see Shafik and Bandyopadhyay [1992], Panayotou [1995, 1997], Shafik [1994], Selden and Song [1994], Torras and Boyce [1996], Cole et al. [1997], de Bruyn et al. [1998], Kaufmann et al. [1998], and Stern et al. [1998] for EKC studies using sulfur. The majority of these studies use a quadratic or cubic specification of income per capita and test whether a significant turning point exists. The estimated threshold levels are substantially different across these studies ranging from $2,894 (Panayotou, 1995) to $12,346 (Kaufmann et al. 1998). This large variation may be attributable to the differences in the source of data, inclusion of additional variables into the model, the use of emission or concentration of sulfur. Usually panel data from the Global Environmental Monitoring System's (GEMS) tracking of urban air quality in different cities in the developing and developed world has been used (Grossman and Krueger 1991, 1995, Panayotou 1997, Shafik 1994, Torras and Boyce 1996); OECD data was the next most commonly used data set (Cole et al. 1997, Selden and Song 1994). Technology level (Cole et al. 1997), locational dummies (Grossman and Krueger 1991, 1995, Shafik 1994), population density (Grossman and Krueger 1991, 1995, Panayotou 1997, Selden and Song 1994), GDP/area, exports/GDP (Kaufmann et al. 1998) were among the additional variables included in the models. A concise tabular summary of studies on sulfur dioxide emissions can be found in Stern et al. [1998].

All these studies have been subject to a modelling criticism: It is unclear why the specific reduced-form equation (the quadratic or cubic specification of pollution with respect to income per capita) employed in their estimations exists. Schmalensee et al. [1998] study uses spline (piecewise linear) estimation, and thus avoids this modelling problem; however, they do not provide any theoretical reasoning for the division of the data into income segments except that each segment contains equal number of observations. Moreover, there is no statistical test on the existence and significance of the threshold. Their approach takes threshold level as pre-determined (the income segments are chosen by the authors) instead of finding it from the *data*. We will address these critical issues in our paper.

One way to overcome the modelling problem is to employ non-parametric kernel estimation method, which does not impose any a priori functional relationship between variables. The other advantage of the nonparametric kernel estimation technique in the present context is that it enables to compute the impact of income or other included variables on the pollution level for *each observation point* in the data set. Thus, the functional behavior of pollution with respect to income can be analyzed without relying upon a priori imposition of quadratic or cubic specifications.

An important feature of our empirical approach, beside the use of nonparametric kernel method, is to test the EKC hypothesis across different parts of a city, namely residential, commercial and industrial areas (which are further divided into centre-of-the-city and suburban areas). This decomposition is justified for the following reasons. Firstly, environmental stresses in a given city will likely be different across residential, commercial and industrial areas. Main sources of pollution are normally in the industrial areas. Secondly, there are asymmetries related to the introduction of new environmental regulations across these areas. Given that economic development is almost the top priority policy item in most of the countries, policy makers give special attention to the competitiveness of their industries, and thus, they tend to soften the level of regulations on the industry. On the other hand, new environmental regulations in the residential areas (such as the type of coal that could be used for heating etc.) are more easily introduced. Thirdly, the policy makers have to face strong industry lobby during both the introduction and implementation of environmental regulations, whereas such a pressure is not present in residential and perhaps in commercial areas. Given this background, it is important that the impact of economic development on the environment should be analyzed across different parts of a city. To our knowledge, no analysis of the EKC hypothesis with this type of decomposition at the city level has been reported before.

Furthermore, the division of data across different parts of city also serves another purpose. The data employed in the previous EKC studies is limited either because of small number of observations or because of limited variation in the income variable in the large samples. That is, the data set includes pollution values from different observation stations in the same city and/or in different cities in a given country. Thus, although pollution shows large variations across these monitoring stations, income per capita, which is defined at the country level, remains constant. This generates poor econometric estimates for the relationship between income and pollution. By separating the estimation procedures across different parts of a city, we provide a partial[1] remedy to this problem.

Finally, we also identify the impact of coastal location on the EKC-type behavior for sulfur dioxide emissions. This is important as climatic knowledge indicates that sea breeze and windy conditions of the coastal cities do not let air pollution to stay for long.

In section 2, we briefly present a discussion of non-parametric kernel estimation technique employed in this paper. Section 3 presents our data sources and model. In section 4, discussion of results is given, and section 5 summarizes main findings.

2. Nonparametric Kernel Estimation

Consider the stochastic process $\{y_t, x_t\}$, $t = 1,2,...,n$; where y_t is a scalar and $x_t = (x_{t1}, x_{t2}, ..., x_{tq})$ is $(1 \times q)$ vector which may contain the lagged values of y_t. The regression model is $y_t = m(x_t) + u_t$, where $m(x_t) = E(y_t \mid x_t)$ is the true but unknown regression function, and u_t is the error term such that $E(u_t \mid x_t) = 0$.

---

[1] This is a partial solution as our data set includes several cities from a given country (47 cities in 28 countries), which will have different pollution levels but same income per capita. A complete solution will require a better data set.

If $m(x_t)$ is a correctly specified family of parametric regression, then one can construct the ordinary least squares (OLS) estimator of $m(x_t)$. For example, if $m(x_t) = a + x_t b = X_t d$, where $d = (a \quad b')'$ and $X_t = (1 \quad x_t)$, is linear we can obtain the OLS estimator of d by minimizing $\sum u_t^2 = \sum (y_t - X_t d)^2$ as

$$\hat{d} = (X'X)^{-1} X'y\,. \tag{2.1}$$

However, it is well known that if the specified regression $X_t d$ is incorrect then the OLS estimates $\hat{d}$, and hence $\hat{m}_t = X_t \hat{d}$ are inconsistent and biased, and they may generate misleading results.

An alternative approach is to use the consistent nonparametric regression estimation of the unknown $m(x)$ by the local linear least squares (LLLS) method. For obtaining the LLLS estimator we first write first-order Taylor series expansion of $m(x_t)$ around x so that

$$\begin{aligned} y_t &= m(x_t) + u_t = m(x) + (x_t - x)m^{(1)}(x) + v_t \\ &= a(x) + x_t b(x) + v_t = X_t d(x) + v_t\,, \end{aligned} \tag{2.2.}$$

where $a(x) = m(x) - xb(x)$, $d(x) = [a(x) b(x)']'$, and $b(x) = m^{(1)}(x)$, and $m^{(1)}$ shows the first derivative. Then, solving the problem:

$$\min \sum_{t=1}^{n} v_t^2 K_{tx} = \min \sum_{t=1}^{n} (y_t - X_t d(x))^2 K_{tx} \tag{2.3}$$

with respect to $d(x)$, we get the LLLS estimator as:

$$\tilde{d}(x) = (X' K(x) X)^{-1} X' K(x) y \tag{2.4}$$

where $K(x)$ is a diagonal matrix of the kernel (weight) $K_{tx} = K((x_t - x)/h)$ and h is the window width. The LLLS estimators of $a(x)$, $b(x)$ and $m(x)$ are calculated as $\tilde{a}(x) = [1 \quad 0]\tilde{d}(x)$, $\tilde{b}(x) = [0 \quad 1]\tilde{d}(x)$ and $\tilde{m}(x) = \tilde{a}(x) + x\tilde{b}(x)$. These LLLS estimators are consistent; for further details on properties, see Fan and Gijbels [1996] and Pagan and Ullah [1999].

The LLLS estimators of $d(x)$ and $m(x)$ are also called the nonparametric kernel estimators, which are essentially the local linear fits to the data corresponding to the x_i's which are in the interval of length h around x, the point at which d is calculated. In this sense the LLLS estimator provides the varying estimates of d with changing values of x. It depends on the kernel function K and the window width h. The function K is chosen to be a decreasing function of the distances of the regressor x_t from the point x, and the window width h determines how rapidly the weights decrease as the distance of x_t from x increases. In our empirical analysis we have considered an optimal parabolic kernel and the cross validated window width; for further details, one can see Pagan and Ullah [1999, ch.3] and Racine [1999].

3. Data and Model

We have used the same GEMS data as in Grossman and Krueger [1995]. The panel data from the Global Environmental Monitoring System's (GEMS) tracking of urban air quality in different cities in the developing and developed world runs from 1977 upto 1991 with a total number of 1478 observations. The data includes 47 cities in 28 countries in 1977, 52 cities in 32 countries in 1982, and 27 cities in 14 countries in 1988; overall, 42 countries are represented in the sample. As expected in such a large data set, there are some missing observations corresponding to either pollution levels or some explanatory variables, and thus, usable number of observations in our case was 1321 (see Table 1). GEMS data is an extensive data and it includes many different pollutants such as sulfur dioxide emissions, smoke, suspended particulates etc. In this paper, we will use the sulfur emissions data.

The generic model employed in the EKC literature is as follows:

$$Y_{it} = G_{it}\, b_1 + G^2_{it}\, b_2 + G^3_{it}\, b_3 + A_{it}\, b_4 + A^2_{it}\, b_5 + A^3_{it}\, b_6 + X_{it}\, b_7 + E_{it}$$

where Y_{it} is a measure of water or air pollution in station i in year t, G_{it} is GDP per capita in year t in the country in which station i is located, A_{it} is the average GDP per capita over the prior three years, X_{it} is a vector of other covariates (like temperature, population density, location dummies such as residential, industrial and commercial), and E_{it} is the error term. This is the model used by Grosmann and Krueger [1995] study. In such a model, the test of EKC hypothesis is done by checking the sign and the significance of coeffcients of GDP per capita terms.

Our estimation methodology differs with Grossman and Krueger [1995] in two key respects. First, we employ non-parametric kernel estimation technique to establish the nature of income-pollution relationship without imposing any a priori restriction on the functional form. Second, we divide our sample into sub-samples, based on the characteristics of the site where the monitoring stations in a given city are located. We first divide the sample into three broad categories, namely residential, industrial and commercial areas, and then whether the site is in the centre-of-the-city or in the suburban area. In this way we classified the sample into nine sub-samples, and for each sub-sample the following model has been estimated:

$$SE = f(GDPPC, PDENS, COAST) \qquad (3.1)$$

where, SE is sulfur dioxide (SO_2) emissions, GDPPC is the GDP per capita, PDENS is the population density, COAST is a dummy indicating whether the city is located along a coastline. It should be noted that no special form has been assumed for the function f(.) in (3.1) in the estimation stage. Equation (3.1) has been estimated for each of the subsamples separately. Details regarding the number of observations in each subsample are given in Table I in the appendix.

In our model, EKC hypothesis will be tested as follows. By estimating equation (3.1) with nonparametric kernel method, we will obtain the gradients (partial derivatives) of SE (sulfur emissions) with respect to each independent variable separately for each observation point. Then, we will plot the gradients of SE with respect to GDP per capita against GDP per capita (by using only the significant estimates). A plot with positive gradients upto a certain income level and with negative ones after that level present support for EKC hypothesis. We now summarize our findings.

4. Discussion of Results

In the first stage, model in (3.1) has been estimated for the three basic sub-samples, residential, industrial and commercial areas. We have obtained the nonparametric LLLS estimates of the gradients, $\tilde{b}(x)$, defined in (2.4), for each regressor at each sample point, with their associated standard errors. In our empirical analysis, we have considered an optimal parabolic kernel and the cross-validated window width (for further details, one see Pagan and Ullah [1999, ch.3]). Econometric estimation of $\tilde{b}(x)$ have been done by employing N© BETA, Computer Software (Racine 1999).

According to our specification in equation (3.1), a derivative estimate $\tilde{b}(x)$ represents the partial derivative of SO_2 emissions with respect to per capita income, population density and coastal-city dummy. Since the focal point of this paper is to examine the EKC hypothesis, we present the results of our varying point estimates with respect to per capita income only. As our estimation technique presents gradients with their standard errors for *each* observation point, we can not list them here due to space constraints[2]. Instead, we will use plots to report our estimation results.

In figure 1 in the appendix, we have plotted the significant gradients with respect to per capita income for the three sub-samples (Residential, Industrial and Commercial) against per capita income. For each group, we have also plotted the observations that correspond to the off-coastal and coastal cities separately in the same figure.

Under EKC hypothesis, one would expect that these gradients would be positive for the low and middle-income economies, and negative for the higher income groups.

Overall these results do not seem to support the hypothesis; however, we do see some evidence of inverted U-type relationship for the observations that correspond to the residential and commercial areas belonging to off-coastal cities.

In the second stage, we further divided these three sub-samples into those that correspond to the centre-of-the-city, and the ones that correspond to the suburban areas, and re-estimated our model in (3.1). The gradients with respect to per capita income have been reported in Figures 2 and 3. In the same figures, we have also plotted observations relating to coastal and off-coastal cities as before separately. In all three cases (Residential and Centre-of-the-City, Industrial and Centre-of-the-City, Commercial and Centre-of-the-City), there is clear evidence that EKC holds (Figure 2). In fact for the Residential and Centre-of-the-City case it holds vividly. The turning points of the inverted U-type relationship vary across different areas. For the residential and commercial areas it is around \$2500, and for the industrial areas, it is around \$5000. In Figure 3, where results for the suburban areas have been plotted, we do not see any indications of EKC type behavior.

Finally, related to the impact of population density variable on the level of sulfur emissions, we have mixed results. For the residential locations (both for the centre-of-the-city and suburban), after a population density of 50,000, the gradient is positive suggesting pollution increases as density increases as reported in earlier studies in the literature. For the industrial and commercial cases, we have mixed results.

In short our results do provide a qualified support for the EKC hypothesis. That EKC hypothesis holds universally can not be supported; it depends on the type of the land use nearby the monitoring station.

---

[2] At least 162 gradients have been estimated in each subset. The authors can provide these estimation results freely if requested.

5. Conclusions

Interactions between economic development and environmental degradation have been investigated intensively. Although no firm conclusions have been achieved unequivocally regarding the impact of higher economic development on the level of different pollutants, an inverted U-type relationship, known as Environmental Kuznets Curve hypothesis, has been commonly proposed especially for sulfur dioxide emissions. Empirical studies indicating evidences for EKC have been subject to an important modelling criticism because different polynomial forms (usually quadratic or cubic) between pollution and income have been estimated without any justification. This paper tests the existence of EKC hypothesis by incorporating nonparametric kernel estimation method. This method does not require any a priori functional specification between dependent and independent variables, and thus, enables us to avoid the modelling problem.

More importantly, we present a test of EKC behavior across different parts of city, namely residential, industrial and commercial. Our results suggest a qualified support for the validity of EKC hypothesis. We show that statistical evidence supporting EKC is location specific. It holds for observations corresponding to the centre-of-the-city parts (whether they belong to residential, commercial or industrial areas). On the other hand, a weak evidence exists only for residential and commercial areas when centre-of-the-city and suburban parts are combined. Furthermore, in the centre-of-the-city cases, where EKC type behavior is observed, the turning point is at higher incomes in the industrial areas as compared to the residential and commercial ones. This is not surprising as growth is high priority in developing countries and thus, severe restrictions on the industry come only in the later stages of the development.

Additionally, a distinction between coastal and off-coastal cities seems to be important as the impact of income on pollution in coastal cities remains negative most of the time. It has been suggested in the literature that sea breeze and windy conditions of the coastal cities do not let air pollution to stay for long, and this may explain why some support for the hypothesis is only evident for the off coastal-city observations.

Our findings indicating a location-specific EKC hypothesis may lead to new research in the EKC literature. Why does the impact of economic development on the environment differ across different parts of city? At a first glance, this can be explained by differences with respect to environmental pressures, and demand for better environment across different areas in a given city. During different phases of economic development, the increased load of city-centres originating from different environmental stress factors, such as traffic, makes centre-of-the-city more vulnerable to, especially, air pollution. As a result of higher income levels, the demand for environmental quality in polluted areas becomes more pronounced, which may lead to the introduction of new regulations. Furthermore, a common observation is that policy makers tend to put more emphasis on infrastructural and environmental improvements in their (big) cities as a sign of “modernization”. Nevertheless, these are only some very initial conjectures on our findings indicating location-specific EKC behavior. Further work is suggested to examine the reasons why EKC holds clearly in the centre-of-the-city areas (whether residential, industry or commercial), and lack of evidence in the other cases. This will require a data set with more details at the micro level. Such an analysis at a more disaggregated level will also be useful in understanding the fundamental linkages between income and environmental quality: whether the transformation into improved environmental conditions is an automatic outcome of higher economic development, or whether it requires a careful coordination of development and environmental policies. We plan to take up these important issues in our future work.

References

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

De Bruyn S. M., J. C. J. M. Van Den Bergh, and J. B. Opschoor, 1998, Economic Growth and Emissions: Reconsidering the Empirical Basis of Environmental Kuznets Curves, *Ecological Economics*, Special Issue on EKC.

Fan, J. and I. Gibels, 1996, Local Polynomial Modeling and Its Applications, Chapman and Hall, London.

Grossman, G.M. and Krueger, A.B., 1991, Environmental Impacts of North American Free Trade Agreement, *NBER Working Paper Series*, No: 3914.

Grossman, G.M., and A.B. Krueger, 1995, “Economic Growth and the Environment”, *Quarterly Journal of Economics*, 110, 353-77

Kaufmann, R.K., Davidsdottir B., Garnham S., Pauly P., 1998, “The Determinants of Atmospheric SO_2 concentrations: reconsidering the environmental Kuznets curve”, *Ecological Economics*, 25, 209-230.

Pagan, A., and A.Ullah, 1999, *Nonparametric Econometrics*, Cambridge University Press, New York and Melbourne.

Panayotou T., 1995, Environmental Degradation At Different Stages Of Economic Development, in: I. Ahmed and J. A. Doeleman (eds.) *Beyond Rio: The Environmental Crisis and Sustainable Livelihoods in the Third World*, MacMillan, London.

Panayotou T., 1997, Demystifying The Environmental Kuznets Curve: Turning A Black Box Into A Policy Tool , *Environment and Development Econ.* 2(4), 465-484.

Racine, Jeff, 1999, N© BETA, Computer Software, University of South Florida.

Schmalensee, R., Stoker, T.M., and Judson, R.A., 1998, “World Carbon Dioxide Emissions: 1950-2050,” *Review of Economics and Statistics* 80(1), 15-27.

Selden, T.M. 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.

Shafik, N. and S.Bandyopadhyay (1992), “Economic Growth and Environmental Quality: Time Series and Cross-Country Evidence”, World Bank Policy Research Working Paper, WPS 904, Washington DC: World Bank.

Shafik N., 1994, “Economic Development and Environmental Quality: An Econometric Analysis”, *Oxford Economic Papers,* 46, 757-773.

Stern, D.I., T. Auld, M.S. Common, and K. Sanyal, “Is there an Environmental Kuznets Curve for Sulfur?” Working Papers in Ecological Economics, No: 9804, Australian National University, (1998).

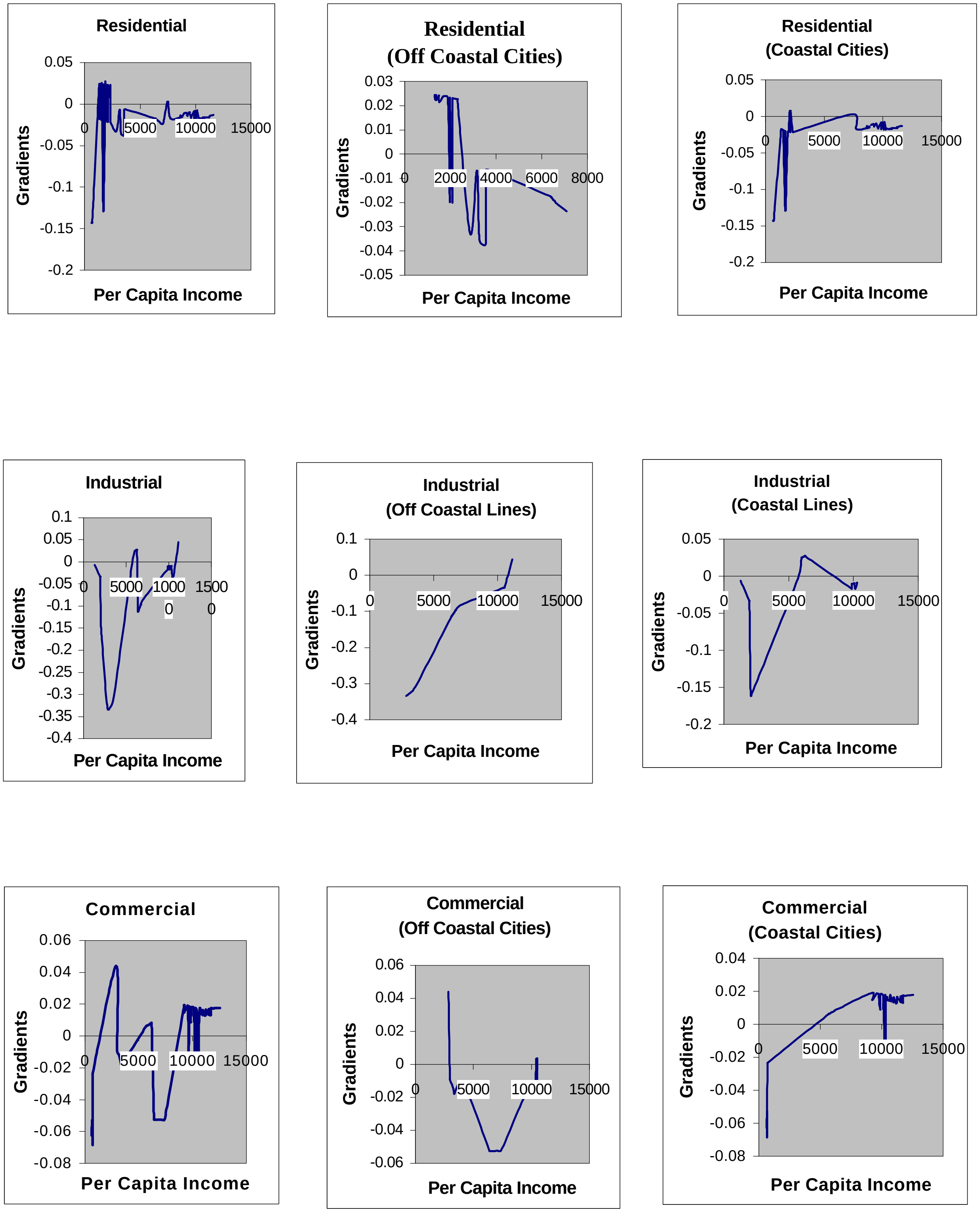
Torras M. and J. K. Boyce, 1996, Income, inequality, and pollution: a reassessment of the environmental Kuznets curve, *Ecological Economics*, special issue on EKC.

Appendix

Table I. Sample sizes.

	Total	Centre of City	Suburban
Residential	487	181	306
Industrial	391	162	229
Commercial	433	397	36
Total	1321	750	571

Figure 1. Emission(SE)-Income(GDPPC) relationship (Gradients, dSE/dGDPPC) across residential, industrial and commercial areas.


Residential

Residential (Off Coastal Cities)

Residential (Coastal Cities)

Industrial

Industrial (Off Coastal Lines)

Industrial (Coastal Lines)

Commercial

Commercial (Off Coastal Cities)

Commercial (Coastal Cities)

Figure 2. Emission(SE)-Income(GDPPC) relationship (Gradients, dSE/dGDPPC) for the center-of-city areas.

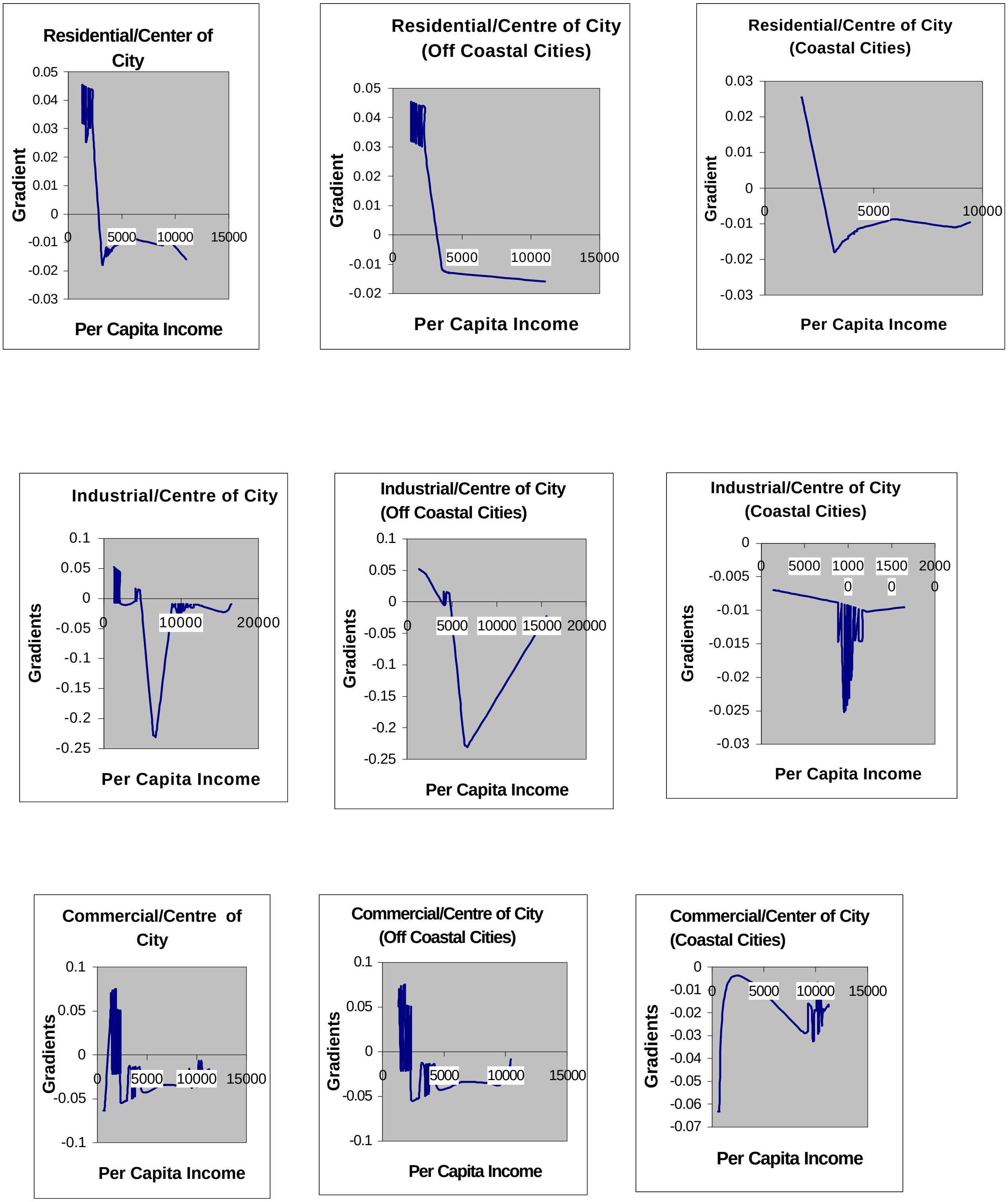


Figure 3. Emission(SE)-Income(GDPPC) relationship (Gradients, dSE/dGDPPC) for the suburban areas (Due to lack of enough number of observations, estimation for commercial and suburban subset could not be done).