Evaluating the Environmental Kuznets Curve

Michael S. Delgado
Department of Economics
Binghamton University

Kai Sun
Department of Economics
Binghamton University

Subal C. Kumbhakar*
Department of Economics
Binghamton University

August 27, 2010

Abstract

We examine the empirical evidence for an environmental Kuznets curve using a semiparametric smooth coefficient regression model that allows us to incorporate flexibility in the parameter estimates, while maintaining the basic econometric structure that is typically used to estimate the pollution-income relationship. This allows us to assess the sensitivity to parameter heterogeneity of typical parametric models used to estimate the relationship between pollution and income, as well as identify why the results from such models are seldom found to be robust. Our results confirm that the resulting relationship between pollution and income is fragile; we show that the estimated pollution-income relationship depends substantially on the heterogeneity of the slope coefficients and the parameter values at which the relationship is evaluated. Different sets of parameters obtained from the semiparametric model give rise to many different shapes for the pollution-income relationship that are commonly found in the literature.

1 Introduction

In a seminal article, Grossman and Krueger (1995) identified an inverse U-shaped relationship between several different measures of aggregate pollution and national income, providing evidence that pollution levels might initially increase as countries develop, but ultimately decline as countries achieve higher levels of income. This phenomenon has been studied extensively (e.g., World Bank (1992), Selden and Song (1994), Grossman and Krueger (1993, 1995), Schmalensee *et al.* (1998), Millimet *et al.* (2003)), and is typically referred to as the environmental Kuznets curve (EKC).

The empirical evidence in favor of the existence of an EKC has been mixed, however. Harbaugh *et al.* (2002) show that cleaning the data and increasing the total number of observations used by Grossman and Krueger (1995) results in a failure to identify the EKC relationship observed in the initial study. While Harbaugh *et al.* (2002) show that the observed EKC relationship is sensitive to the included sample of observations, their primary results assume the same parametric specification used by Grossman and Krueger (1995).[1]

Using the data from Harbaugh *et al.* (2002), we extend their study by incorporating flexibility in the parameter estimates through the use of semiparametric kernel methods. The advantage of our

*Correspondence to: Subal C. Kumbhakar, Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902. Phone: (607) 777-4762, Email: kkar@binghamton.edu

[1]Harbaugh *et al.* (2002) also estimated the model using a spline regression, the results of which are consistent with those from the parametric model.

approach is that we obtain a distribution of coefficient estimates through which we can evaluate the pollution-income relationship, while maintaining their basic econometric framework. Our results show that the existence of the EKC relationship identified by Grossman and Krueger (1995) does not necessarily disappear after polishing and expanding the data as Harbaugh *et al.* (2002) claim. Rather, the existence of the EKC depends on the parameter values used to estimate the pollution-income relationship; our results give rise to sets of parameters that generate the pollution-income paths identified by both Grossman and Krueger (1995) and Harbaugh *et al.* (2002).

We point out that this paper is not the first to apply nonparametric methods to the empirical EKC literature; see, for example, Millimet *et al.* (2003), Bertinelli and Strobl (2005), Paudel *et al.* (2005), Azomahou *et al.* (2006). Our approach, however, differs in the following ways. First, each of these papers incorporates country (or site) and time fixed effects but assume the effects enter into the specification linearly and additively separable from the unknown regression function. While this is often more convenient for nonparametric estimation of the unknown regression function, this fails to allow the partial effects to vary with respect to the country or time effects, nor does it allow for interaction between the fixed effects and other control variables. Our approach allows the fixed effects to enter into the unknown coefficient function and interact with other regressors, providing coefficient estimates that vary with respect to site and time specific effects.

Second, our approach allows us to obtain a distribution of estimated intercept and GDP slope coefficients with which to evaluate the pollution-income relationship. The advantage of this is that we can test the sensitivity of the EKC shape to the parameter values at which the relationship is evaluated without changing the underlying econometric framework. Typical parametric models assume coefficients are constant across observations; our results show this assumption has strong implications for both the robustness and shape of the pollution-income relationship.

Third, the smooth coefficient regression framework allows us to maintain a basic quadratic (or cubic) relationship between pollution and GDP while still allowing for flexibility in the functional form of the regression model and obtain observation specific coefficients. Previous studies applying nonparametric techniques completely abandon the commonly used quadratic (or cubic) framework; the problem with abandoning the basic parametric structure is that it is difficult to isolate which assumptions inherent in the parametric model are necessary for identifying an inverted U-shaped relationship between pollution and GDP. Our approach allows us to directly examine both the effectiveness and weakness of traditional parametric models in estimating the pollution-income relationship.

2 Nonparametric Estimation

Both Grossman and Krueger (1995) and Harbaugh *et al.* (2002) estimate a reduced form regression model in which the dependent variable, the mean level of pollution, depends on both gross domestic product and a three-year average of lagged GDP. The mean level of pollution is assumed to be a cubic function of both GDP and lagged GDP. We exclude the three-year average of lagged GDP from our primary regressions since parametric results with or without lagged GDP are qualitatively identical (Grossman and Krueger (1995)).[2] Additionally, we exclude the third order polynomial of GDP since a cubic specification is unreliable when estimating the regression function at high levels of income (Grossman and Krueger (1995)).[3]

We generalize the regression specification by applying the semiparametric smooth coefficient model proposed by Li *et al.* (2002). The smooth coefficient model is given by

$$y_i = \beta_0(z_i) + \beta_1(z_i)GDP_i + \beta_2(z_i)GDP_i^2 + \epsilon_i \qquad i = 1, 2, \ldots, n \tag{1}$$

where y_i is the mean level of SO_2 at the i^{th} monitoring site, z_i is a vector of exogenous variables for which the coefficients of GDP are allowed to vary, and $\beta_0(z_i)$, $\beta_1(z_i)$, and $\beta_2(z_i)$ are unknown, smooth functions of z_i. Specifically, z_i includes national population density, the share of investment in physical capital to GDP, trade intensity, an index of democratization, national GDP relative to the average GDP, and categorical indicators controlling for monitoring site and time effects. One advantage of the smooth coefficient model is that it nests the standard constant parameter model as a special case. That is, when the coefficients are constant ($\beta_j(z_i) = \beta_j \ \forall \ j$) the model is fully parametric and can be estimated using OLS. Other parametric functional forms on $\beta_j(z_i)$ can also be specified (which will also make the model fully parametric) and these parametric functions can be tested econometrically against the smooth coefficient model in (1) using the test proposed by Li *et al.* (2002).

There remains one final econometric issue that needs attention prior to discussing the results, namely, how to assess the relationship between pollution and GDP in the presence of additional regressors and parameter heterogeneity. Harbaugh *et al.* (2002) and Grossman and Krueger (1995) set all remaining variables at their mean level, and allow them to enter the regression function as a constant.[4] This implicitly assumes all countries in the sample can be adequately represented by the mean level of each regressor. While this assumption is arguably innocuous in the parametric model, since the mean level is absorbed by the intercept term and will not affect the slope of the pollution-income relationship, it is potentially quite restrictive in the semiparametric model in which the coefficients are allowed to vary across each observation. We, therefore, check the robustness of the pollution-income relationship by evaluating the relationship at different percentiles of the estimated coefficients. If the relationship is markedly different when evaluated at different percentiles, we can be fairly certain that observations with qualitatively different slope coefficients should not, at least for this sample of observations, be evaluated at one particular observation. This complication makes accurate identification of the pollution-income relationship difficult, and demonstrates why the results from parametric models are often brittle when finding evidence in favor of or against the

[2]Results from the semiparametric model that includes lagged GDP yield qualitatively identical results to the primary results reported in this paper.

[3]Additional results that include the 3rd order polynomial of GDP yield estimates consistent with our primary results. Moreover, there are many empirical studies on the EKC that include only a second order polynomial of GDP. See, for example, Selden and Song (1994), Cole *et al.* (1997), Stern and Common (2001), Khanna (2002).

[4]Additionally, Grossman and Krueger (1995) evaluated GDP based on the sum of the coefficients for GDP and lagged GDP. Our omission of lagged GDP allows us to avoid this complication.

EKC hypothesis.

3 Results

We estimate four different models, each following the structure given in (1). These models differ in terms of the z_i variables in the $\beta(\cdot)$ coefficients. The $\beta(\cdot)$ coefficients in Model 1 are functions of national population density. In Model 2, the coefficients are functions of national population density and a categorical indicator for monitoring site, and in Model 3, the coefficients are functions of national population density and physical capital's share of GDP. The coefficients in Model 4 are functions of all of the z_i variables mentioned in section 2. With regards to correct model specification, we implement the specification test proposed by Li *et al.* (2002), and for each model we reject the null hypothesis that the coefficients are some parametric function of the variables in z_i, at the 1% significance level. These results are consistent with those of Millimet *et al.* (2003), who find evidence in favor of a semiparametric model compared to standard parametric alternatives. For details regarding the test statistic and testing procedure, see Li *et al.* (2002).

In order to identify an inverted U-shaped relationship, our parameter estimates must satisfy the following conditions. First, we require the coefficient on the second order polynomial of GDP to be negative, that is, $\beta_2(z_i) < 0$, at least for the coefficient value used to evaluate the pollution-income relationship. This condition is sufficient to ensure the relationship between pollution and GDP is concave. The necessary condition to obtain an economically meaningful peak (peak for a positive level of GDP) in the level of pollution is for the coefficient on the first order polynomial of GDP to be positive, that is, $\beta_1(z_i) > 0$, at the coefficient value used to identify the pollution-income relationship. Since the semiparametric model provides different slope coefficients for each observation in the sample, we can use the above conditions to identify whether or not the EKC relationship is consistent across observations.

Figure 1 shows a scatter plot of the estimated slope coefficients for each of the four models. In order to satisfy the necessary and sufficient conditions mentioned above, we require the estimated coefficients to reside in the fourth quadrant (bottom right) for each model. The figure shows that the conditions are jointly satisfied for only a subset of the observations in the sample. Specifically, the conditions are satisfied for 53%, 45%, 45%, and 49% of the observations for Models 1 through 4, respectively. Thus, the pollution-income relationship is consistent with the EKC hypothesis for approximately half of the observations in the sample, and the identification of the relationship will be sensitive to the coefficients at which the relationship is evaluated.[5]

Although Figure 1 clearly shows that the EKC relationship does not hold for all data points, we now follow the standard practice and evaluate the pollution-income relationship, for each of the estimated models, using different evaluation points. Table 1 displays a summary of the distribution of estimated coefficients and their corresponding wild bootstrapped standard errors for Model 1.

[5] The scatter plot does not show the EKC relationship directly because the plot is for $\hat{\beta}_1(z_i)$ and $\hat{\beta}_2(z_i)$. However, since the necessary and sufficient conditions for an EKC relationship is expressed in terms of signs on $\hat{\beta}_1(z_i)$ and $\hat{\beta}_2(z_i)$, this plot is very useful as it shows the violation points.

We identify an inverted U-shaped relationship between pollution and GDP at the median coefficient values, while the previously mentioned conditions are violated at the 25^{th} and 75^{th} percentiles. If we evaluate the pollution-income relationship using the median coefficient values, we are able to identify a peak in the level of pollution at approximately \$4,000. The first panel of Figure 2 shows a plot of this relationship.

Tables 2 through 4 contain a summary of the results from Models 2 through 4. We find that in each case, the coefficient restrictions necessary for identifying an inverted U-shaped relationship are violated at each reported percentile. The second, third, and fourth panels of Figure 2 contain plots of the pollution-income relationship for each the models (Models 1-4), in which the coefficients are evaluated at the 25^{th}, 50^{th}, and 75^{th} percentiles. We find that the coefficients across each model, evaluated at the 25^{th} and 75^{th} percentiles, all follow similar patterns, regardless of which z_i variables are included in the model. This can be clearly seen in the second and fourth panels. While we fail to identify an inverted U-shaped relationship between pollution and GDP for any of these models, our estimated pollution-income relationship is substantially different from those estimated by Grossman and Krueger (1995) and most of those estimated by Harbaugh *et al.* (2002). This suggests that the pollution-income relationship is heavily dependent on the sample of observations, the coefficient evaluation point, and the model specification. Many of the estimated relationships identified by Harbaugh *et al.* (2002), as well as by Grossman and Krueger (1995), ultimately curve up or down at higher income levels, an artifact of the cubic specification used in both studies (see Grossman and Krueger (1995) for a brief discussion).

For the models evaluated at the median coefficient values, we find that the pollution-income relationship varies greatly with the included variables in the model, identifying an EKC relationship only for the model that includes national population density. This does not mean that the EKC relationship holds for every observation in Model 1, as shown in the first panel of Figure 1. Thus, although there may not be any EKC relationship (controlling for other variables at their actual/observed values), a spurious relationship might emerge when the relationship is evaluated at certain percentiles. In addition, we find that when including monitoring site fixed effects to the original model that identifies the inverted U-shaped relationship, the EKC relationship disappears and becomes linear (see Models 1 and 2 in Figure 2). This potentially suggests that any models controlling for fixed effects effectively removes site-specific (or country-specific) information, that is crucial for correctly identifying the pollution-income relationship across countries.

4 Conclusion

We examine the empirical evidence for an environmental Kuznets curve using flexible econometric techniques that are capable of incorporating heterogeneity in the slope coefficients. Our results show that the existence of an EKC relationship is fragile, and dependent on heterogeneity of the slope coefficients and the coefficient values at which the pollution-income relationship is evaluated. In addition, we find that the pollution-income relationship is sensitive to the functional form assumptions of the model; our results show how the different shapes of the pollution-income relationship identified by previous studies are consistent with the distribution of parameter values estimated using the semiparametric model.

Acknowledgements

We thank Bill Harbaugh for providing us with the data. Daniel Henderson provided useful comments. All nonparametric estimation was conducted using version 0.30-7 of the "np" package (Hayfield and Racine (2008)) in R version 2.10.1.

References

[1] Azomahou, T., F. Laisney, and P. N. Van, 2006. "Economic Development and CO2 Emissions: A Nonparametric Panel Approach," *Journal of Public Economics*, 90, 1347-1363.

[2] 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.

[3] Grossman, G. M. and A. B. Krueger, 1993. "Environmental Impacts of a North American Free Trade Agreement," *The U.S.-Mexico Free Trade Agreement*, P. Garber, ed. MIT Press, Cambridge, MA.

[4] Grossman, G. M. and A. B. Krueger, 1995. "Economic Growth and the Environment," *Quarterly Journal of Economics*, 110 (2), 353-377.

[5] Harbaugh, W. T., A. Levinson and D. M. Wilson, 2002. "Reexamining the Empirical Evidence for an Environmental Kuznets Curve," *Review of Economics and Statistics*, 84 (3), 541-551.

[6] Hayfield, T. and J. S. Racine, 2008. "Nonparametric Econometrics: The np Package," *Journal of Statistical Software*, 27. URL http://www.jstatsoft.org/v27/i05/.

[7] Khanna, N. 2002. "The Income Elasticity of Non-point Source Air Pollutants: Revisiting the Environmental Kuznets Curve," *Economics Letters*, 77, 387-392.

[8] Li, Q., C. J. Huang, D. Li, and T. Fu, 2002. "Semiparametric Smooth Coefficient Models," *Journal of Business and Economic Statistics*, 20 (3), 412-422.

[9] Millimet, D. L., J. A. List, and T. Stengos, 2003. "The Environmental Kuznets Curve: Real Progress or Misspecified Models?" *Review of Economics and Statistics*, 85 (4), 1038-1047.

[10] R Development Core Team, 2009. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

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

[12] 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.

[13] Stern, D. I. and M. S. Common, 2001. "Is There an Environmental Kuznets Curve for Sulfur?" *Journal of Environmental Economics and Management*, 41, 162-178.

[14] World Bank, 1992. *World Development Report 1992: Development and the Environment*, Washington, D.C.

Table 1: Summary of estimated smooth coefficients for Model 1.

Coefficient	Q1	Q2	Q3
$\beta_0(z_i)$	11.73	48.57	105.6
	(0.61)	(0.81)	(6.14)
$\beta_1(z_i)$	-12.67	0.96	17.25
	(0.82)	(0.19)	(0.45)
$\beta_2(z_i)$	-1.63	-0.13	0.67
	(0.04)	(0.02)	(0.07)

Notes: Wild bootstrapped standard errors are in parentheses. Coefficients are functions of national population density, and Q1, Q2, and Q3 refer to the 25^{th}, 50^{th}, and 75^{th} percentiles, respectively.

Table 2: Summary of estimated smooth coefficients for Model 2.

Coefficient	Q1	Q2	Q3
$\beta_0(z_i)$	22.47	58.53	148.00
	(2.64)	(0.83)	(5.45)
$\beta_1(z_i)$	-28.01	-1.12	11.73
	(0.89)	(0.24)	(1.13)
$\beta_2(z_i)$	-0.85	0.00	1.57
	(0.05)	(0.01)	(0.10)

Notes: Wild bootstrapped standard errors are in parentheses. Coefficients are functions of national population density and a categorical monitoring site indicator. Q1, Q2, and Q3 refer to the 25^{th}, 50^{th}, and 75^{th} percentiles, respectively.

Table 3: Summary of estimated smooth coefficients for Model 3.

Coefficient	Q1	Q2	Q3
$\beta_0(z_i)$	-6.79	66.06	118.90
	(1.70)	(0.96)	(5.48)
$\beta_1(z_i)$	-19.40	-0.75	21.65
	(2.53)	(0.18)	(0.58)
$\beta_2(z_i)$	-1.20	-0.10	1.12
	(0.05)	(0.01)	(0.18)

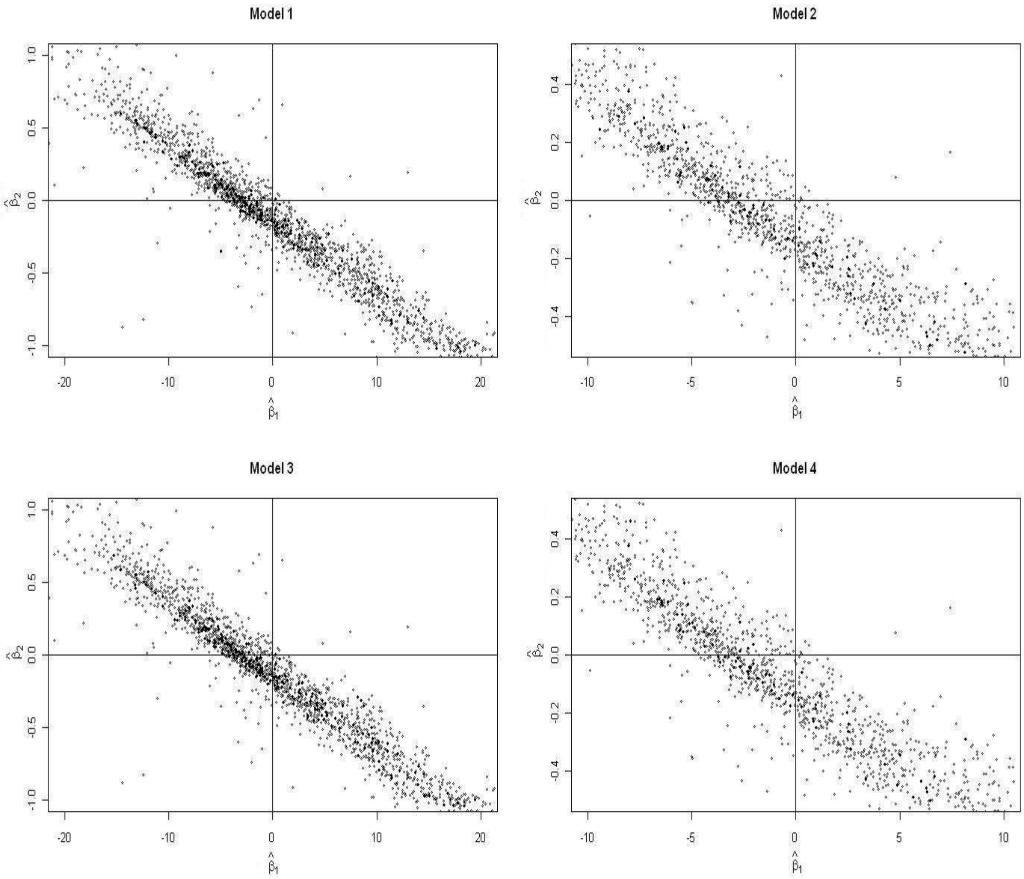
Notes: Wild bootstrapped standard errors are in parentheses. Coefficients are functions of national population density and physical capital investment. Q1, Q2, and Q3 refer to the 25^{th}, 50^{th}, and 75^{th} percentiles, respectively.

Table 4: Summary of estimated smooth coefficients for Model 4.

Coefficient	Q1	Q2	Q3
$\beta_0(z_i)$	35.52	64.29	83.20
	(0.37)	(0.76)	(7.72)
$\beta_1(z_i)$	-6.58	-0.16	9.62
	(0.26)	(0.21)	(0.25)
$\beta_2(z_i)$	-0.64	-0.14	0.20
	(0.01)	(0.01)	(0.01)

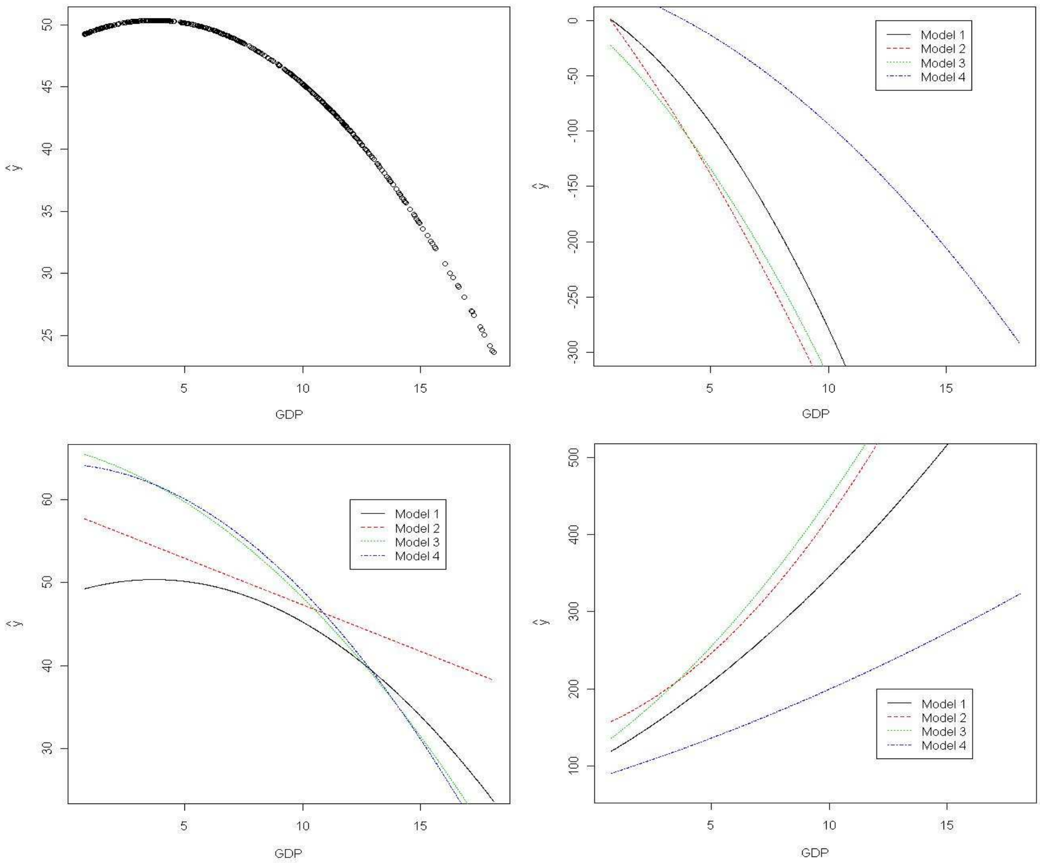
Notes: Wild bootstrapped standard errors are in parentheses. Coefficients are functions of national population density, physical capital investment, trade, democracy, relative GDP, and categorical indicators for monitoring site and time. Q1, Q2, and Q3 refer to the 25^{th}, 50^{th}, and 75^{th} percentiles, respectively.

Figure 1: Scatter plot of the estimated coefficients for Models 1-4.



Notes: Points shown in the fourth quadrant (bottom right) of each panel are consistent with an inverted U-shaped pollution-income relationship. All other points contain violations in the conditions necessary to identify an EKC.

Figure 2: Plot of the pollution-income relationship across Models 1-4.



Notes: The top left panel shows the pollution-income relationship for Model 1, evaluated at the median coefficient value. The other panels show the pollution-income relationship for each model, evaluated at the 25^{th} (top right), 50^{th} (bottom left), and 75^{th} (bottom right) percentiles.