CrossMark

ORIGINAL PAPER

Revisiting the Environmental Kuznets Curve: a semi-parametric analysis on the role of market structure on environmental pollution

Michael L. Polemis[1,2]

Abstract The scope of this paper is to assess the validity of the Environmental Kuznets Curve hypothesis by focusing for the first time in the literature on the impact of market structure on industrial pollution proxied by the level of toxic chemical releases. For this reason, we used a flexible semiparametric fixed effects regression estimator in the spirit of Baltagi and Li (Ann Econ Finance 3(1):103–116, 2002). The empirical analysis is based on a panel data set including industrial facilities for the US states over the 1987–2012 period. Contrary to the parametric results, we uncover an inverted "*U-shaped*" relationship between industrial output and toxic chemical releases when we account for the presence of market concentration.

Keywords Market structure · Industrial pollution · Semiparametric fixed effects model · Non-linearities · EKC hypothesis

JEL Classification Q52 · L1 · C14

1 Introduction

Grossman and Krueger (1995) introduced the Environmental Kuznets Curve (EKC) concept as a way to describe correlations in aggregate data that might subsequently be explained through more detailed analysis involving transmission pathways and micro

✉ Michael L. Polemis
  mpolemis@unipi.gr

1   Department of Economics Piraeus, University of Piraeus, Piraeus, Greece

2   Hellenic Competition Commission, Athens, Greece

⚛ Springer

foundations.[1] This has generated an expansive literature that treats the EKC model itself as providing a causal theory. However, EKC models typically omit variables such as energy prices and policy instruments (i.e financial variables, etc) that are known to causally influence pollutants like CO_2 and SO_2 emissions (Polemis and Stengos 2017).

The majority of the EKC studies estimate reduced-form equations that enter the model in a parametric form with controversial results (Hsueh 2015; Figueroa and Pasten 2013; Halkos and Tzeremes 2009). Parametric estimates assume a unique response coefficient for the covariates in environmental pollution regressions. However, this assumption is not warranted since in the EKC literature, theory and empirical evidence emphasize the existence of multiple regimes (i.e threshold effects) followed by strong-non linear effects (Halkos and Polemis 2016; Millimet et al. 2003). Following this spirit we apply a non-parametric technique to extend this line of research. In this way, we are able to trace possible non-linear effects in more detail and accuracy and shed some light to the competition-pollution nexus totally ignored by the existing literature.

Despite the profound interest by policy makers and government officials on the possible spillovers between market concentration and environmental degradation the existing literature is still in its infancy, with controversial results. These can be justified by the fact that many researchers acknowledge that competition may have positive as well as negative effects on environmental pollution (Simon and Prince 2016; Branco and Villas-Boas 2015; Shleifer 2004).

We contribute to this strand of literature by applying a flexible Semi-Parametric Fixed Effects Model (SPFEM) to analyze the impact of industrial output on environmental degradation when we account for the level of market concentration developed by Baltagi and Li (2002) with significant applications (see among others Libois and Verardi 2013; Desbordes and Verardi 2012). The main advantage of this technique is that it allows us to investigate any possible spillover effect between market structure and industrial pollution without knowing either the form of this relationship (i.e curvature) or the exact interactions among the main variables of interest. In other words, one of the possible gains from using non-parametric techniques is that they do not rely on strong assumptions of the functional form, usually without an initial analysis on the properties of the data. It is worth mentioning that such an ad hoc analysis using parametric methods can lead to biased estimation of economic relationships (Polemis and Stengos 2015; Delis et al. 2014; Tran and Tsionas 2010).

Our findings reveal the existence of an inverted "U-*shaped*" relationship between industrial output and environmental degradation already hidden when we impose parametric techniques. The rest of the paper is organized as follows. Section 2 describes the data. Section 3 develops the econometric methodology and the SPFEM, while Sect. 4 focuses on the empirical findings. Lastly, Sect. 5 concludes the paper.

---

[1] EKC hypothesis implies a non linear relationship of an inverted '*U*' type between environmental degradation and economic growth.

2 Data

Toxic chemical releases were drawn from the Toxic Release Inventory of Environmental Protection Agency covering the period 1987–2015. The panels used in this study consist of 2461 industrial facilities broken down by 473 6-digit NAICS codes over 5 year intervals; namely 1987, 1992, 1997, 2002, 2007 and 2012.[2]

The sample period is selected in order to reduce the measurement error bias generated by the use of higher-frequency data (Griliches and Hausman 1986). The structural variables such as market concentration and value added that correspond to each 6-digit code were drawn from the National Bureau of Economic Research (NBER) and especially from Manufacturing Industry Database (CES).

This database contains annual industry-level data from 1958 to 2011 on output, employment, payroll and other input costs, investment, capital stocks, and various industry-specific price indexes. Because of the change from SIC to NAICS industry definitions in 1997, the database is provided in two versions (one with 459 four-digit 1987 SIC industries and the other with 473 six-digit 1997 NAICS industries). Especially for the year 2012, and due to data restrictions concerning the level of market concentration as measured by certain indicators (i.e CR4, CR8, CR20, CR50 and HHI), we used data directly from the US Census of Manufacturers. The latter is only conducted every 5 years limiting our time span to 6 years (1987, 1992, 1997, 2002, 2007 and 2012). Similarly to Polemis and Stengos (2017) and in order to check for the robustness of our findings, we take five measures of market concentration: HHI is the Herfindahl-Hirschman index for the 50 largest firms in the industry[3], CR4 is the four-firm concentration ratio, CR8 is the eight-firm concentration ratio, CR20, is the twenty-firm concentration ratio and finally CR50, is the fifty-firm concentration ratio.[4] It is worth mentioning that our measures of market structure reveal the existence or the absence of effective competition in the industry since concentration is simply the inverse of competition (Simon and Prince 2016).

3 Empirical framework

We estimate the SPFEM following the methodology described in Baltagi and Li (2002). Let the model be given by the following equation:

$$y_{it} = a_i + x_{it}^T \beta + w_{it} \gamma + f(z_{it}) + \varepsilon_{it} \tag{1}$$

where $f(z_{it})$ is an unknown function of z_{it}, entering the model in a non linear way. Y_{it} is the dependent variable. X_{it} is the vector of exogenous linear regressors,

---

[2] The sample period is selected in order to reduce the measurement error bias generated by the use of higher frequency data (Griliches and Hausman 1986).

[3] The HHI ranges from zero (Perfect competition) to unity (Monopoly).

[4] To preserve space we limit our analysis to CR4 and HHI concentration measures. The rest set of the results regarding the other market structure indicators (CR8, CR20 and CR50) are available upon author's request.

while the w-vector includes the year dummy variables. Lastly, ε_{it} are zero mean i.i.d. innovations.

Following Baltagi and Li (2002), we approximate $f(z_{it})$ by series differences $p^K(z_{it})$ where the latter are the first k terms of a sequence of functions $[p_1(z), p_2(z), \ldots]$. By taking first differences in order to remove fixed effects, we end up with the following equation:

$$\Delta(y_{it}) = \Delta(x_{it}^T)\beta + \Delta(w_{it})\gamma + \Delta\{p^k(z_{it})\}\delta + \Delta(\varepsilon_{it}) \tag{2}$$

Equation 2 can be estimated by using OLS. In the next step, we use the fitted fixed effects $\hat{a}_i$ in order to estimate the error component residual of Eq. 1. Thus we have:

$$\hat{u}_{it} = y_{it} - x_{it}^T\hat{\beta} - w_{it}\hat{\gamma} - \hat{a}_i = f(z_{it}) + \varepsilon_{it} \tag{3}$$

As it is evident with the above transformation we eliminate the fitted fixed effects $\hat{a}_i$ (Baltagi and Li 2002). Therefore, in the next step, we could estimate f(z_{it}) using a nonparametric estimator based on a kernel local polynomial fit or spline interpolation. We use the latter approach (B-spline of order K = 2) since it better approximates complex shapes and does not suffer from Runge's phenomenon (Newson 2012).

4 Results and discussion

We carry out the first part of the analysis with the existence of cross-section independence since it is common for panel data to violate this assumption which will result in low power and size distortions of tests. We use the cross-section dependence test proposed by Pesaran (2004). The test strongly rejects the null hypothesis of cross-section independence for all the variables, providing evidence of cross-sectional dependence in the data given the statistical significance of the cross-section dependence (CD) statistic (Table 1).

In light of this evidence we proceed to test for unit roots using tests that are robust to cross-section dependence ("*second generation*" tests) proposed by Breitung and

Table 1 Pesaran CD test

Variable	Value	p value	Correlation
ln(REL)	133.05***	0.000	0.692
ln(SHIP)	143.34***	0.000	0.768
ln(SHIP)^2	143.34***	0.000	0.768
ln(CR4)	157.56***	0.000	0.845
ln(HHI)	103.39***	0.000	0.596

Under the null hypothesis of cross-section independence the CD statistic is distributed as a two-tailed standard normal. Results are based on the test of Pesaran (2004). The p values are for a one-sided test based on the normal distribution. Correlation denotes the average value of the off-diagonal elements of the cross-sectional correlation matrix of residuals obtained from estimating Eq. 3. Significant at ***1%

Table 2 Unit root tests

Variable	Breitung and Das (2005)	Pesaran (2007)
ln(REL)	− 23.0000***	− 51.4745***
ln(SHIP)	− 23.5125***	− 46.2411***
ln(SHIP)^2	− 23.6784***	− 44.4356***
ln(CR4)	− 37.8671***	− 54.1824***
ln(HHI)	− 39.6548***	− 53.0001***

The null hypothesis assumes that the variable contains unit root. The number of lags has been set to two according to BIC. Significant at ***1%

Das (2005) and Pesaran (2007). Both tests suggest that all the sample variables are stationary (Table 2).

In the next step we estimate separately (see Eqs. 4, 5) the following (parametric) panel data models by employing various techniques such as pooled OLS, Fixed Effects (P-FE) and Instrumental Variables fixed effects (IV P-FE) with the latter controlling for possible endogeneity and reverse causality generated by the inclusion of industry output (SHIP) in the relevant regressions. For this reason we followed a similar approach with the study of Halkos and Polemis (2017) in which a simple model is developed that incorporates the relationship of financial development and economic growth to environmental degradation together with the validation of the EKC hypothesis. Our parametric equations thus take the following form:

$$\ln(REL_{jit}) = \alpha_i + \beta_t + b_0 + b_1 \ln(SHIP_{it}) + b_2 \ln(SHIP_{it})^2$$
$$+ b_3 \ln(HHI_{it}) + u_{jit} \tag{4}$$
$$\ln(REL_{jit}) = \alpha_i + \beta_t + b_0 + b_1 \ln(SHIP_{it}) + b_2 \ln(SHIP_{it})^2$$
$$+ b_3 \ln(CR4_{it}) + u_{jit} \tag{5}$$

Where $\ln(REL_{ijt})$, denotes the logged total chemical releases emitted by facility j in industry i across the year t. $Ln(SHIP_{it})$ is the logged value of shipments as a proxy for industry output i during year t. $Ln(HHI_{it})$ is the logged Herfindahl-Hirschman index for the 50 largest firms in the industry i, while $\ln(CR4_{it})$ is the four-firm concentration ratio in the industry i expressed in natural logarithm. Moreover, α_i is the unit-specific residual that differs between sectors but remains constant for any particular facility (unobserved facility level effect) while β_t captures the time effect and therefore differs across years but is constant for all sectors in a particular year. Finally, u_{jit} denotes the idiosyncratic error term which is i.i.d. From Table 3, it is shown that all the coefficients of the SHIP terms are statistically significant alternating their signs starting from negative to positive, while the two market concentration indices are positively correlated with the chemical releases.[5] In other words, industry output decreases up to a certain "*turning*" point and then increases gradually. This suggests the existence of a '*U*' shaped relationship between industrial output and pollution when market concentration is taken into consideration. However, the Wooldridge test (W-T) for serial correlation in the error term reveals that there is a lot of serial dependence in

---

[5] If $b_1 < 0$, and $b_2 > 0$, then we come up with a "*U-shaped*" quadratic relationship.

Table 3 Estimation results

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled OLS	Pooled OLS	P-FE	P- FE	IV P-FE	IV P-FE
$\ln$(SHIP)	−19.63** (8.482)	−14.75* (7.873)	−18.60** (8.611)	−13.74*(7.975)	−28.80*** (8.619)	−26.89*** (8.484)
$\ln$(SHIP)^2	10.05** (4.241)	7.559* (3.936)	9.537** (4.306)	7.054* (3.988)	14.61*** (4.309)	13.67*** (4.242)
$\ln$(HHI)	0.429*** (0.0221)	–	0.503*** (0.0548)	–	0.410***(0.0222)	–
$\ln$(CR4)	–	0.824*** (0.0355)	–	0.827*** (0.0945)	–	0.781*** (0.0396)
Constant	0.674* (0.346)	2.119*** (0.280)	−0.00517 (0.351)	1.340*** (0.280)	0.853** (0.347)	0.267 (0.352)
Diagnostics						
Observations	11,729	14,264	11,729	14,264	11,706	11,706
R-squared	0.0751	0.0716	0.075	0.0722	0.0756	0.0769
Wald /F statistic	1,154.64*** [0.000]	1,429.31*** [0.000]	119.24*** [0.000]	138.57*** [0.000]	490.32*** [0.000]	539.97*** [0.000]
W-T	364.48*** [0.000]	931.61*** [0.000]	362.18*** [0.000]	214.42*** [0.000]	256.56*** [0.000]	345.67*** [0.000]
HW	–	–	27.72*** [0.000]	51.29*** [0.000]	–	–

Robust standard errors are in parentheses. Time effects are estimated but not reported. Fixed effects are justified by the Hausman test. The numbers in square brackets are the p values. P-FE, Parametric Fixed Effects model; IV P-FE, Instrumental Variables Parametric Fixed Effects model. W-T is the Wooldridge test for serial correlation in the error term. HW is the Huber/White test for groupwise heteroscedasticity. Significant at ***1%, **5% and *10% respectively
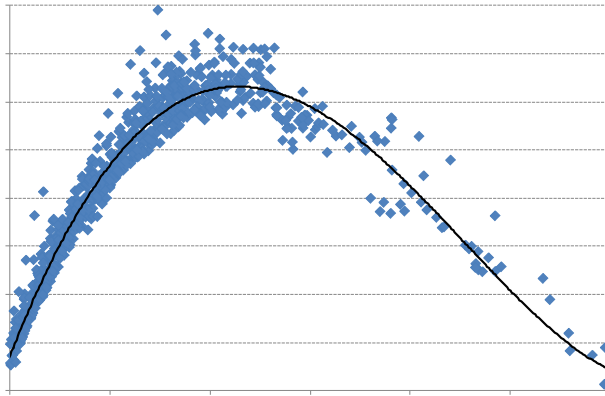
Fig. 1 Inverted U-shape curve between industrial output and pollution. The horizontal axis measures the logged industrial output ln(SHIP), while in the vertical axis the logged level of industrial toxic releases ln(REL) is measured. The points in the graph are partial residuals for industrial toxic releases which are centered around the mean. Blue shaded area corresponds to 95% confidence bands

the residuals and the standard errors since the null hypothesis (homoskedasticity) is strongly rejected in all of the six alternative specifications with a p value equal to 0.000.[6]

In the next step, we proceed to estimate the SPFEM given by the following equation:

$$\ln(REL_{jit}) = \alpha_i + \beta_t + f[\ln(SHIP_{it}) + \ln(SHIP_{it})^2] + b_1 \ln(HHI_{it}) + u_{jit} \quad (6)$$

where f(·) is a function which is non-parametrically estimated. The graphical presentation of the semi-parametric estimation of f(·) is displayed in Fig. 1.[7] As it is evident, the SPFEM redefines the curvature as the relationship between industrial output and pollution when market concentration is taken into account appears to be inversely "*U-shaped*" and is now precisely estimated. In other words, competition intensity tends to decrease the level of toxic chemical releases and thus industrial pollution up to a certain point reversing its trend henceforth.

These results highlight the superiority of semi-parametric models in exploring non-linear relationships compared to classical parametric approaches. This could be attributed to the fact that parametric regression models are highly nonlocal as the conditional expectation in one part of the data can be heavily influenced by observations in another part (Desbordes and Verardi 2012).

5 Conclusions

In this paper we compare and contrast standard parametric techniques with a non-parametric one in order to study the impact of market concentration on environmental

---

[6] We greatly thank an anonymous referee for raising this issue.

[7] To preserve space we present the curvature of the semi-parametric model with the Hirschman-Herfindahl index.

pollution within an EKC framework. In this way we draw sharp differences between the usual parametric methodologies (i.e Pooled OLS, FE-OLS and IV) and the semiparametric fixed effects estimator along the lines of Baltagi and Li (2002).

The empirical findings justify for the first time in the empirical literature the existence of an inverted "U-*shaped*" curve linking toxic chemical pollution with industrial output when market concentration is taken into account. This relationship provides new insights into the environmental policy since the regulators must take into account if they are on the upward or the downward slopping part of the curve in order to pursue the effective environmental policies. Lastly, we argue that using a flexible semi-parametric fixed effects estimator we do not misleadingly fail to find an inverted U-shape curve linking competition with environmental degradation compared with the parametric techniques.

The existence of an (inverted) "U" shaped curve provides new insights into the environmental policy debate toward emissions releases abatement. This means that the increasing non parametric regression line up to a certain concentration level approximately indicates a negative effect on facilities' emissions levels whereas a decreasing line indicates a positive effect. Lastly our semiparametric econometric model concurs that the results remained robust under different specifications of market concentration.

References

Baltagi, B.H., Li, D.: Series estimation of partially linear panel data models with fixed effects. Ann. Econ. Finance **3**(1), 103–116 (2002)

Branco, F., Villas-Boas, J.: Competitive vices. J. Mark. Res. **52**, 801–816 (2015)

Breitung, J., Das, S.: Panel unit root tests under cross-sectional dependence. Stat. Neerl. **59**(4), 414–433 (2005)

Desbordes, R., Verardi, V.: Refitting the Kuznets curve. Econ. Lett. **116**, 258–261 (2012)

Delis, M., Iosifidi, M., Tsionas, E.G.: On the estimation of marginal cost. Oper. Res. **62**(3), 543–556 (2014)

Figueroa, E., Pasten, R.: A tale of two elasticities: a general theoretical framework for the environmental Kuznets curve analysis. Econ. Lett. **119**(1), 85–88 (2013)

Griliches, Z., Hausman, J.A.: Errors in variables in panel data. J Econom. **31**(1), 93–118 (1986)

Grossman, G.M., Krueger, A.B.: Economic growth and the environment. Quart. J. Econ. **110**(2), 353–377 (1995)

Halkos, G.E., Polemis, M.L.: Does financial development affect environmental degradation? Business strategy and the environment, evidence from the OECD Countries (2017). https://doi.org/10.1002/bse.1976

Halkos, G., Polemis, M. (2016). The good, the bad and the ugly? Balancing Environmental and Economic Impacts Towards Efficiency. MPRA Paper 72132, University Library of Munich, Germany

Halkos, G., Tzeremes, N.G.: Exploring the existence of Kuznets curve in countries' environmental efficiency using DEA window analysis. Ecol. Econ. **68**, 2168–2176 (2009)

Hsueh, L.: Regulatory effectiveness and the long-run policy horizon: the case of U.S. toxic chemical use. Environ. Sci. Policy **52**, 6–22 (2015)

Libois, F., Verardi, V.: Semiparametric fixed-effects estimator. Stata J. **13**(2), 329–336 (2013)

Millimet, D., List, J., Stengos, T.: The environmental Kuznets curve: real progress or misspecified models? Rev. Econom. Stat. **85**(4), 1038–1047 (2003)

Newson, R.: Sensible parameters for univariate and multivariate splines. Stata J. **12–3**, 479–504 (2012)

Pesaran, M.H. (2004). General diagnostic tests for cross section dependence in panels. Cambridge Working Papers in Economics 0435, Faculty of Economics, University of Cambridge

Pesaran, M.: A simple panel unit root test in the presence of cross-section dependence. J. Appl. Econom. **22**, 265–312 (2007)

Polemis, M., Stengos, T.: Does competition prevent industrial pollution? Evidence from a Panel Threshold Model, working Paper Series 17-07, The Rimini Centre for Economic Analysis (2017)

Polemis, M., Stengos, T.: Does market structure affect labour productivity and wages? Evidence from a smooth coefficient semiparametric panel model. Econ. Lett **137**, 182–186 (2015)

Shleifer, A.: Does competition destroy ethical behavior? Am. Econ. Rev. **94**, 414–418 (2004)

Simon, D.H., Prince, J.T.: The effect of competition on toxic pollution releases. J. Environ. Econ. Manag. (2016). https://doi.org/10.1016/j.jeem.2016.03.001

Tran, K., Tsionas, E.G.: Local GMM estimation of semiparametric panel data with smooth coefficient models. Econom. Rev. **29**, 39–61 (2010)