ECONOMICS

Working Paper Series

No. 01-2024

Does green innovation crowd out other innovation of firms? - based on the extended CDM model and unconditional quantile regressions

Richard S.J. Tol
Department of Economics, University of Sussex, Falmer, BN1 9SL, UK
Institute for Environmental Studies, Vrije Universiteit, Amsterdam, The Netherlands
Department of Spatial Economics, Vrije Universiteit, Amsterdam, The Netherlands
Tinbergen Institute, Amsterdam, The Netherlands
CESifo, Munich, Germany
Payne Institute for Public Policy, Colorado School of Mines, Golden, CO, USA
College of Business, Abu Dhabi University, UAE
R.Tol@sussex.ac.uk

Yi Yiang
Department of Economics, University of Sussex, Falmer, BN1 9SL, UK

Abstract: In the era of sustainability, firms grapple with the decision of how much to invest in green innovation and how it influences their economic trajectory. This study employs the Crepon, Duguet, and Mairesse (CDM) framework to examine the conversion of R&D funds into patents and their impact on productivity, effectively addressing endogeneity by utilizing predicted dependent variables at each stage to exclude unobservable factors. Extending the classical CDM model, this study contrasts green and non-green innovations' economic effects. The results show non-green patents predominantly drive productivity gains, while green patents have a limited impact in non-heavy polluting firms. However, in high-pollution and manufacturing sectors, both innovation types equally enhance productivity. Using unconditional quantile regression, I found green innovation's productivity impact follows an inverse U-- out effects, Productivity, CDM framework, Quantile regression, Recentered influence function

Key words: D24, O31, Q55

2. LITERATURE REVIEW

circumvent endogeneity, simultaneity and reverse causality, thus clarifying the causal links between R&D, innovation outputs, and productivity. Marin et al. (2014) expanded this model, distinguishing between environmental and other forms of innovation to infer crowding-out effects. Applying the classic and extended CDM models, this study analyzes the full sample & K L Q D ¶ V O L V W V X E V D P S O H V W R G L V W L Q J X L V K differentiating high-pollution from non-pollution firms and contrasting high-tech with low-tech, and manufacturing with non-manufacturing sectors.

3.1 R&D equation

In addressing R&D investment decision biases, this study employs a modified Heckman model. In the CDM framework, the Heckman model originally addresses self-selection bias in R&D investment decisions. This study modifies the model to focus on the self-selection bias in the disclosure of R&D expenditure among Chinese listed firms, all of which engage in R&D activities[1]. Approximately 32.9% of firm-year observations between 2010 and 2018 lack R&D expenditure data due to disclosure O L N H O \ L Q I O X H Q F H G E \ I L U P V ¶ U H S X W D W L R confidence. The modified approach uses a probit model for the Inverse Mill's Ratio (IMR) in its first step, estimating the probability of R&D expenditure disclosure based on various controls and exclusion criteria. 7 K H , 0 5 W K H Q F R U U H F W V I R U E L D V L Q W K H V H F assessment. This approach, previously adopted in studies like Edeh and Acedo (2021) and Wang et al. (2021), effectively handles the non-disclosure biases in R&D expenditure. Details of the equations for the IMR calculation and bias correction are provided.

[1] Being large entities, all Chinese listed firms included in this study engage in R&D activities. The absence of R&D expenditure data in certain cases is attributed to non-disclosure rather than a lack of R&D activity, as evidenced by further-disclosed and positive patent data.

R&D intensity. R&D intensity (4 & + 0) is then predicted as logged R&D expenditure per thousand employees, using the Heckman model's estimation.

$$4\&+0_{it}6 = \dot{U}_4 H J 2 2_{it-1} \% + \dot{U}_5 . ' 8_{it} + \dot{U}_6 H J ' / 2_{it} + \dot{U}_7 H J \% \# 2_{it} + \dot{U}_8 H J 2 \%_{it} + 6$$
$$\dot{U}_9 \% 4_{it} + \dot{U}_{10} 5 1 '_{it} + \beta\&_{it} + \acute{e}+ / 4_{it} + \tilde{n}_{it} + \ddot{U}_{it} + \acute{Y}_{it} \quad (4)$$

The environmental policy indicators use the logarithm of lagged provincial pollutant charge (H J 2 2 $\%_{it-1}$), a fee levied on businesses for pollutant emissions, including sewage, waste gas, and hazardous waste[2] (Guo et al, 2019). A higher 2 2 % indicates greater environmental costs for firms, thereby serving as a valid indicator for measuring the crowding-out effects of environmental policy on a firm's R&D investments.

To enhance robustness, the analysis incorporates a two-year lagged PPC (H J 2 2 $\%_{it-2}$) as an alternative proxy, inspired by Yuan and Zhang (2017) who found that a one-year lag in environmental regulations reduced R&D spending, while a two-year lag enhanced it. Additionally, the one-year lagged provincial investment in industrial pollution treatment (H J + 2_{it-1}) is used, which capture the total expenses incurred by industrial firms for pollution treatment at the provincial level. These expenses are funded by pollution charges, government subsidies, and enterprise self-financing. This measure not only reflects the financial burden of environmental protection on firms but also indicates the intensity of local environmental policy implementation.

[2] Under China's 2003 "Measures for the Administration of Pollutant Discharge fees," industrial and commercial entities emitting pollutants are charged a fee, which doubles if emissions exceed national standards. Consequently, a uniform pollution levy rate applies nationwide, meaning a higher provincial pollutant charge indicates greater total emissions within that province.

impact of predicted patent intensity at different points of the productivity distribution without relying on specific conditional assumtions that are required by Conditional Quantile Regression (CQR), preferable when conditional relationships are uncertain and complex.

$$\pm + (\lhd \text{U\'a} R: (\grave{o}; = @ \not{Q}_{\grave{o}} \text{ L } r \qquad : z;$$

$$\pm 4 + (\lhd \text{U\'a} R: (\grave{o}; = @ \not{Q}_{\grave{o}} \text{ L } R: (\grave{o}; \qquad : \{;$$

$$R: (_{\dot{i}} ; 1\ 0\ \text{D} R: (\grave{o}; \acute{a}_{\dot{c}}^{\dot{o}} \cdot \text{E} \qquad : s\ r;$$

$$\hat{e}_{\text{A}\,\dot{c}}^{\delta} \text{ L } \pm + (\lhd \text{U\'a} R: (\grave{o}; = ^{\delta} @ \not{Q}_{\grave{o}} \qquad : s\ s;$$

UQR analyzes how changes in explanatory variables, X, affect the unconditional distribution of the response variable (Rios-Avila, 2020). It is based on the idea of recentered influence functions (RIF) for robust estimation against outliers and understanding the distribution structure. Influence functions (IF) utilizes Gateaux directional derivatives to estimate partial effects, revealing how small changes in distribution affect mean values. RIFs assess the impact of individual observations on the mean, with IFs having an expected value of 0 (Equax8iRJ (r

IF is the influence of an individual observation i on the quantile (q_τ) of unconditional distribution of y; IRIF is the unconditional distribution function of y that is estimated based on the marginal density of y at quantile q_τ that is estimated based on kernal density distribution. To sum up, RIF equals to original distributional statistic at quantile q_τ plus the marginal effects of y made on the quantile q_τ of the distribution.

$$RIF(y;q_\tau) = q_\tau + \frac{\tau - I(y \le q_\tau)}{f_y(q_\tau)} \qquad (13)$$

The coefficient matrix is estimated as the Equation 13, in which it computes sample quantile of marginal distribution and acquire density estimate by kernel density to obtain RIF. The coefficient represents the marginal effect estimates of an infinitesimal location shift in the distribution of covariates X on q_τ (RIF-unconditional quantile regression), ceteris paribus. The study employs individual-time fixed effects RIF regression to examine the heterogeneity in the effects of green and non-green innovations across productivity distribution in the Extended CDM model.

3.3.2 Robustness checks for non-linear relationship estimations

3.3.2.1 RIF regression with treatment effects

To check the robustness of non-linear relationship estimates in the productivity equation, a dummy variable for green patent ownership is used as a treatment variable in RIF regressions

[3] rifhdreg with abs(id year) command allows controlling for high-dimensional fixed effects, which is equivalent to rifhdfe fixed effects settings and estimate within firm effects when controlling individual

with generalized influence treatment effect, identifying effects of investments. However, standard RIF is limited to local approximations, especially for categorical variables. To the confoundedness assumption, it requires to

3.3.2.2 Conditional quantile regression for panel data

Conditional Quantile Regression (CQR), while common in quantile analysis, has limitations due to its reliance on numerous specific covariates, potentially reducing reliability across different quantiles (Bui and Imai, 2018). However, in our research, CQR is effective in CDM frameworks, using patent predictions derived from earlier stages based on limited covariates, then applied in the productivity equation to mitigate

In evaluating the R&D equation, the Heckman selection model, compared with basic OLS, showed no significant multicollinearity (mean VIF = 1.84, see Table B1). The Heckman model's significant lambda (IMR variable) across all samples indicates sample selection bias, effectively addressed by this model (part 3 in Table 1). The first Heckman step revealed environmental performance disclosure (EPD) positively correlated with R&D expenditure disclosure in the full and certain sub-samples, while longer stock market presence (AGE) and larger assets (lnASSET) correlated negatively across all samples, validating their roles as effective exclusion restrictions (part 2 in Table 1).

In the second Heckman step (part 2 in Table 1), environmental regulations (PPF) negatively impacted R&D intensity in all samples but low-tech industries,[7] indicating a crowding-out effect. In the full sample, a 1% rise in pollution charges correlates with a 0.08% decrease in R&D spending, where a provincial increase of 7.7 million in charges typically reduces a firm's R&D investment by 0.4 million RMB. Moreover, non-polluting, high-tech, and non-manufacturing firms showed greater sensitivity (higher elasticity) to environmental costs compared to their counterparts.

Other negative influences on R&D investment included financial risk (LEV), employee costs (EMP), production costs (PCINT) and market concentration (CR4). State-owned enterprises (SOE), despite governmental support, were less efficient in technological innovation. Capital intensity (CAPINT) shows a positive relationship with R&D intensity across most samples, except in high-pollution and manufacturing firms. Bartoloni (2013) highlights the crucial impact of capital structure on R&D investment. In capital-intensive sectors

[7] Low-tech firms, with typically smaller R&D budgets, display negligible changes in R&D investment in response to th environmental policies.

alternative models than non-heavy polluting firms (M3-M6 in Table 2). This disparity is partly due to cost transference within supply chains, with high-polluting firms passing additional environmental expenses onto other businesses. Moreover, the prevailing trend towards environmental responsibility and evolving consumer expectations encourages low-polluting companies to voluntarily augment their environmental investments, thus diverting resources from R&D activities

4.3 Productivity equation

The study employs the Mundlak test over the Hausman test for the productivity equation due to its reliability under issues including heteroskedasticity or serial correlation in error terms (Mundlak, 1978). The Mundlak test, incorporating panel-level means in the random effects model, significantly rejects the null hypothesis at a 99.9% confidence level (Table B7)[8]. This suggests a correlation between unobservable factors in the error term and control variables, indicating the suitability of a panel fixed effects model. The CDM and extended CDM models use individual and time dual fixed effects, with bootstrapping standard errors to address these unobservable factors.

To address the discrepancy between predicted and actual patent counts, and mitigate endogeneity arising from omitted variables like patenting preferences, the model modifies the predicted patent counts[9]. This is achieved by scaling them with each firm's actual average patent number. This adjustment method effectively neutralizes potential biases from unobserved time-invariant factors associated with patenting, while not affecting the variability in firms' productivity, thereby enhancing the accuracy of the productivity equation estimates[10].

Productivity equation models reveal that predicted patent intensity significantly boosts productivity, with a 1% increase in patent intensity leading to a 0.42% increase in productivity in the full sample (Model 43 in Table 5). Capital intensity also positively impacts productivity, while firm size shows negligible effects.

Table 5 Results of productivity equation - CDM

[8] Table A8 also shows Mundlak test results for Extended CDM model. The results also indicate random effects model is biased and fixed effects model should be used.

[9] The exponential transformed green patent and green patent applications compared the realized patent value are presneted in Table B5 and Table B6, respectively

[10] In Marin (2014), the absence of adjustments for potential discrepancies between predicted and actual patent counts may have contributed to their significantly different findings, particularly regarding the crowding-out effects in high-polluting firms.

	Model 43	Model 44	Model 45	Model 46	Model 47	Model 48	Model 49
	Full-sample	High-pollution	Non-pollution	High-tech	Low-tech	Manufacturing	Non-manufacturing
v_r	0.419***	0.469***	0.338***	0.387***	0.423***	0.391***	0.502***

Table7 Results oĎQR on CDM models (Full sample)

Full sample	Model 57	Model 58	Model 59	Model 60	Model 61	Model 62	Model 63	Model 64	Model 65

Contrary to the non-pollution sample where negative coefficients of GI suggest crowding out effects (Table 10), UQR in the high-pollution sample reveals a notably robust impact of GI on productivity, surpassing non-green innovations within the 40th to 80th productivity quantiles (Table 9). This finding underscores the absence of crowding out effects and highlights the beneficial role of GI in enhancing the performance of medium-high productive, high-polluting firms. However, at the 90th productivity quantile, GI's productivity returns diminish, showing lower benefits compared to non-green innovations, indicating crowding-out effects.

In addition to innovations, the impact of different resources on productivity varies across industrial types and productivity levels. Both CDM and extended CDM models show that capital intensity generally boosts productivity more in high-polluting firms (Table C1-Table C2 and Table 9-Table 10). In contrast, non-polluting firms rely more on capital at high productivity (

Figure 5 Non-linear estimations based on UQR of Extended CDM models, high-tech v.s. low-tech sample

Table 11 Results of UQR on Extended CDM models (High-tech sample)

High-tech	Model 129

Differences in crowding-out effects of GI between high-pollution and non-pollution samples aren't mirrored in high-tech vs. low-tech firms (Figure 5). In the extended CDM model, both firm types show similar patterns in non-green and green innovation's impact on productivity, and neither benefits significantly from IGI (

Table 11 and

Table 12). This similarity suggests pollution intensity, not technological level, primarily drives GI's economic returns, making high-tech and low-tech firms suitable for a placebo test to confirm pollution's influence on GI activities.

Table 12 Results of UQR on Extended CDM models (Low-tech sample)

Low-tech	Model 138	Model 139	Model 140	Model 141	Model 142	Model 143	Model 144	Model 145	Model 146
UQR	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
lnNECOINT	0.397***	0.402***	0.328***	0.264***	0.274***	0.298***	0.396***	0.471***	0.604***
	(5.041)	(7.062)	(6.443)	(5.677)	(5.854)	(6.074)	(6.908)	(6.689)	(5.483)
lnECOINT	-0.049	-0.075	-0.022	0.024	0.052	0.032	-0.026	-0.066	-0.074
	(-0.710)	(-1.462)	(-0.480)	(0.566)	(1.233)	(0.742)	(-0.513)	(-1.067)	(-0.782)
lnCAPINT	0.106***	0.126***	0.126***	0.118***	0.103***	0.100***	0.130***	0.126***	0.183***
	(3.993)	(5.997)	(6.531)	(6.512)	(5.800)	(5.266)	(5.695)	(4.627)	(4.696)
lnEMP	-0.051	-0.005	-0.007	-0.011	-0.022	-0.056*	-0.054*	-0.044	-0.058
	(-1.237)	(-0.181)	(-0.260)	(-0.453)	(-0.939)	(-2.329)	(-1.895)	(-1.367)	(-1.198)
_cons	5.298***	5.496***	5.900***	6.296***	6.725***	6.980***	6.993***	7.330***	7.576***
	(23.496)	(32.203)	(38.542)	(44.773)	(47.522)	(45.418)	(38.056)	(33.323)	(23.285)
Individual FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
adj. Rsq	0.644	0.644	0.644	0.644	0.644	0.644	0.644	0.644	0.644
F-Stat	23.74	23.74	23.74	23.74	23.74	23.74	23.74	23.74	23.74

_cons	5.369***	5.507***	5.865***	6.455***	6.669***	7.000***	7.222***	7.536***	7.991***
	(26.030)	(31.796)	(37.001)	(45.726)	(47.381)	(44.111)	(38.995)	(35.648)	(24.098)
Individual FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

The study's insights are crucial for corporate strategy, highlighting that the relative benefits

R I * , Y D U \ Z L pollution level and productivity level. For high-

	(3.082)	(3.728)	(3.859)	(1.966)	(3.400)	(3.830)	(2.341)	(2.015)	(-1.042)
lnEMP	0.075	0.044	0.045	-0.081**	-0.071**	-0.154***	-0.118***	-0.134**	-0.315***
	(1.476)	(1.111)	(1.106)	(-2.648)	(-2.757)	(-4.509)	(-3.614)	(-3.234)	(-3.527)
_cons	5.717***	6.122***	6.410***	7.133***	6.977***	6.838***	7.165***	7.576***	9.072***
	(24.387)	(28.377)	(24.231)	(38.550)	(45.062)	(41.631)	(40.106)	(33.806)	(19.199)
Individual FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
adj. Rsq	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668
F-Stat	20.63	20.63	20.63	20.63	20.63	20.63	20.63	20.63	20.63
Prob > F	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Log likelihood	-5059	-5059	-5059	-5059	-5059	-5059	-5059	-5059	-5059
N	4,594	4,594	4,594	4,594	4,594	4,594	4,594	4,594	4,594

Standard errors in parenthesis

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

The RIF treatment effects models, both standard and IPW-adjusted, demonstrate that & K L Q D ¶ V Oi nLvHVWLQJ LQ GUHHQ LQQRYDWLRQ positively exhibit higher productivity up to the 80th productivity quantile compared to those without (Figure 9, Table 15 and Table D.1). While UQR analysis previously reveals some negative, albeit insignificant, marginal effects of GI, RIF treatment effects affirm the positive average economic impact of GI, particularly at lower levels of productivity.

Figure 10 indicates that in high-polluting firms, GI's impact on productivity shifts with productivity levels. Firms with lower productivity initially face negative effects, hindered by limited capabilities in efficient technology implementation. As productivity increases, these firms become more adept at leveraging new technology, both green and non-green innovations,

core product innovations predominate and the gains from resource efficient GI become more limited.

Figure 11 Non-linear estimations based on RIF treatment effects (pollution sample)

For non-polluting firms, GI yields positive economic returns for firms at lower productivity levels, peaking at the 20th productivity quantile. The RIF treatment effects models, both standard and IPW-adjusted, demonstrate that & K L Q D ¶ V Oi Lv Vi Wh Ug i G Hl Co Ri Ev Wh Ly consistently exhibit higher productivity up to the 80th productivity quantile compared to those without GI (Figure 9, Table 15 and Table D1). While UQR analysis previously reveals some negative, albeit insignificant, marginal effects of GI, RIF treatment effects affirm the positive average economic impact of GI particularly at lower levels of productivity.

Figure 10 indicates that in high-polluting firms, GI's impact on productivity shifts with productivity levels. Firms with lower productivity initially face negative effects, hindered by limited capabilities in efficient technology implementation. As productivity increases, these firms become more adept at leveraging new technology, both green and non-green innovations, for productivity gains (Table 16). This highlights the critical role of a firm's 00009(icie)is hie00000912000

ECO	0.200*	1.111***	0.909***	0.727***	0.544***	0.371***	0.185***	-0.140**	-0.612***
	(2.221)	(42.774)	(33.409)	(26.205)	(19.514)	(12.699)	(5.564)	(-3.262)	(-11.855)
lnNECOINT	0.423**	0.147***	0.177***	0.197***	0.183***	0.201***	0.252***	0.274***	0.314***
	(2.950)	(7.282)	(8.190)	(9.143)	(8.896)	(8.895)	(8.926)	(7.267)	(5.673)
lnCAPINT	0.318	0.01	0.023	0.025	0.037*	0.02	0.007		

full, non-pollution, high-tech, low-tech, and non-manufacturing samples indicate crowding-out effects of GI in firms without environmental issues.

5. CONCLUSION

The article begins by estimating the classical CDM model to examine the economic effects of R&D activities in listed companies. It then introduces an extended CDM model to test for a crowding-out effect of green innovation (GI) on other innovations, by assessing the separate impacts of GI and non-green innovations on firm productivity. Additionally, unconditional quantile regression (UQR) is employed to investigate potential non-linear relationships between innovations and productivity across the productivity distribution, and to examine how crowding-out effects may vary among different types of firms at different levels of productivity. Finally, the robustness checks conduct conditional quantile regression (CQR) estimations as well as recentered influence function (RIF) estimations that compare the treatment effects on productivity when a firm invests in GI compared to non-GI investments.

The results of the first step of the CDM model, which estimates R&D intensity, indicate that environmental regulations impose additional cost burdens and reduce funds available for firms' R&D investments. The analysis corrects for firm selection bias in R&D expenditure disclosure using the Heckman model. In the second step, the bias-corrected R&D intensity is used to predict firms' patents in the CDM model and green/non-green patents in the extended CDM models. The findings suggest that the efficiency of R&D input-output conversion is low for GI in firms without pollution issues but high in firms with pollution issues, particularly high-polluting firms and manufacturing firms. This implies that crowding-out effects of GI in terms of

Lastly, the estimations of recentered influence function (RIF) treatment effects indicate that, on average, having GI brings positive productivity growth below the 80th quantile of the productivity distribution. However, it becomes pure costs for firms without pollution issues at the top-level of productivity. High-polluting firms with productivity levels above the median are more likely to benefit from the reduction of pollution costs or efficiency improvements induced by GI, as there may be a capability threshold for transforming technology into productivity.

Despite the valuable insights provided by this research, there are some limitations to be considered. Firstly, the analysis is based on data from listed companies, which may not fully represent the entire population of firms. The findings may not be generalized to non-listed or smaller firms that might have different characteristics and resource constraints. Secondly, the study focuses on the impact of GI on firm productivity, neglecting other important factors that could influence productivity, such as market competition and managerial practices. Future research should consider exploring the mechanisms through which GI impacts productivity, such as the role of organizational capabilities, knowledge transfer, and technological spillovers, could deepen our understanding of the underlying processes driving the relationship. Moreover, future studies could investigate the long-term effects of GI on firm performance and sustainability, and

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used Chat-GPT in order to improve the language. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

Barbieri, N., Ghisetti, C., Gilli, M., Marin, G., & Nicolli, F. (2017). A survey of the literature on environmental innovation based on main path analysis. Environmental Economics and Sustainability,

221-250.

Bartoloni, E. (2013). Capital

Marin G. (2014).Do eco-innovations harm productivity growth through crowding out? Results of an extended CDM model for Italy. Research Policy.Mar 1;43(2):301-17.

Marin G, Lotti F. (2017).Productivity effects of eco-innovations using data on eco-patents. Industrial and corporate change. Feb 1;26(1):125-48.

Mundlak, Y. (1978): On the pooling of time series and cross section data. Econometrica 46:69-85.

Norton-Griffiths, M., & Southey, C. (1995). The opportunity costs of biodiversity conservation in Kenya. Ecological economics, 12(2), 125-139.

Popp D, Newell RG. (2009) Where does energy R&D come from? Examining crowding out from environmentally-friendly R&D. National Bureau of Economic Research;Oct 13.

3 R U W H U 0 (D Q G / L Q G H & Y D Q G H U µ 7 R Z D U G D 1 H Z & R Q F H S W L R & R P S H W L W L Y H Q H V V 5 H O D W L R Q V K L S ¶ - R X U Q D O 1 8 R I (F R Q R P L F 3 H U V S H F 10.1257/jep.9.4.97.

Porter ME, Van der Linde C. (1995)Toward a new conception of the environment-competitiveness relationship. Journal of economic perspectives. Dec;9(4):97-118.

Song Y, Yang T, Zhang M. (2019) Research on the impact of environmental regulation on enterprise technology innovation ² an empirical analysis based on Chinese provincial panel data.

Wang Y, Li C, Khan MA, Li N, Yuan R. (2021) Firm information disclosure environment and R&D investment: Evidence from Internet penetration. PloS one. Mar 18;16(3):e0247549.

Yuan B, Xiang Q. (2018) Environmental regulation, industrial innovation and green development of Chinese manufacturing: Based on an extended CDM model. Journal of cleaner production. 2018 Mar 1;176:895-908.

Yuan B, Zhang K. (2017) Can environmental regulation promote industrial innovation and productivity? Based on the strong and weak Porter hypothesis. Chinese Journal of Population Resources and Environment. Oct 2;15(4):322-36.

Zhu, J., Fan, Y., Deng, X., & Xue, L. (2019). Low-carbon innovation induced by emissions trading in China. Nature communications, 10 (1), 1-8.

SOE	15,714	0.46	0.50	0	1

Table A2 Variable explanations

Type	Symbol	Variable	Definition	Source
Dependent variable	RD	R&D dummy	Dummy variable that whether a firm disclose its R&D expenditure information	Wind Database
	RDINT	R&D expenditure intensity	The natural logarithm of the amount of R&D expenditure per employee of listed firms (in millions RMB per thousand employees).	Wind Database
	PAT	Patent applications	Rolling average of firms' patent applications in three years.	Wind/CNRDS
	lnPATINT	Patent intensity	The logarithm of patent applications per employee of listed firms	Wind/CNRDS
	ECO	Eco-patent applications	Rolling average of firms' eco-patent applications in three years.	CNRDS
	lnECOINT	Eco-patent intensity	The logarithm of eco-patent applications per employee of listed firms	CNRDS
	NECO	Other patent applications	Rolling average of firms' patent applications other than eco-patent applications in three years (All patent application minus eco-patent applications).	Self-calculated
	lnNECOINT	Other patent intensity	The logarithm of other patent applications per employee of listed firms	Self-calculated
	lnVA	Value added	The natural logarithm of total sales revenue per employee of listed firms.	CSMAR
Key variables	lnPPC	Provincial pollution charge	The natural logarithm of total amount of pollutant emission charges (in millions RMB).	China Environmental Yearbook
	lnIPT	Industrial pollution treatment investment	The natural logarithm of total amount of industrial pollution treatment investment in each provinces (in millions RMB).	China Statistical Yearbook
	lnFSTK	Firm technological knowledge stock	The natural logarithm of firms' patent applications in the past 2 years	Wind/CNRDS
	lnPTL			

	CR4	Industrial market concentration	The market occupation ratio of the largest four firms in the industry	Self-calculated
Exclusion restrictions				

	(0.018)	(0.018)	(0.046)	(0.045)	(0.043)	(0.043)	(0.021)	(0.021)	(0.022)	(0.022)	(0.039)	(0.039)	(0.023)	(0.023)
lnPPF(2)	0.0325**		0.0209											

Sample	R/P	Obs	Mean	Std.Dev.	Min	Max
Full sample	Realistic	15,714	61.95	293.54	0	10045.00
	Predicted	14,904	63.14	284.35	0	7977.02
High-pollution	Realistic	4,833	45.20	212.67	0	4416.33
	Predicted	4,648	46.12	216.31	0	5294.40
Non-pollution	Realistic	10,881	69.39	322.76	0	10045.00
	Predicted	10,256	70.86	309.35	0	7914.95
High-tech	Realistic	8,361	83.58	340.96	0	10045.00
	Predicted	7,805	86.69	328.02	0	7637.60
Low-tech	Realistic	7,353	37.36	225.47	0	4718.00
	Predicted	7,099	37.25	222.33	0	5801.64
Manufacturing	Realistic	9,207	80.43	328.57	0	10045.00
	Predicted	8,676	82.63	314.92	0	7705.33
Non-manufacturing	Realistic	6,507	35.81	232.76	0	4718.00
	Predicted	6,228	35.99	232.39	0	5838.48

(0.051)

Table C2 Results

Table C4 Results of UQR on CDM models (Low-tech sample)

Low-tech	Model 120	Model 121	Model 122	Model 123	Model 124	Model 125	Model 126	Model 127	Model 128
UQR	Q10	Q20	Q30	Q40	Q50	Q60			

Standard errors in parenthesis
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Figure D 1 Non-linear estimations based on RIF treatment effects (high-tech sample)

	(1.614)	(2.375)	(3.708)	(2.306)	(0.588)	(-0.624)	(-0.982)	(-0.196)	(-0.492)
Yr	0.435***	0.404***	0.414***	0.343***	0.327***	0.348***	0.374***	0.406***	0.504***
lnNECOINT	(8.941)	(9.914)	(10.798)	(9.250)	(9.074)	(9.067)	(9.537)	(8.597)	(7.211)
lnCAPINT	0.063**	0.105***	0.109***	0.092***	0.081***	0.077***	0.101***	0.094***	0.105**
	(2.742)	(5.017)	(5.026)	(4.440)	(4.051)	(3.599)	(4.179)	(3.421)	(2.603)
lnEMP	-0.068*	-0.033	-0.043	-0.055*	-0.060*	-0.062*	-0.052*	-0.067*	-0.081*
	(-1.819)	(-1.068)	(-1.443)	(-1.891)	(-2.182)	(-2.177)	(-1.783)	(-1.984)	(-1.763)
_cons	5.725***								