ARTICLE IN PRESS

Journal of Environmental Economics and Management xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

Is environmental regulation the answer to pollution problems in urbanizing economies?☆

JunJie Wu [a,*], Kathleen Segerson [b], Chunhua Wang [c]

[a] *Department of Agricultural and Resource Economics, North Carolina State University, 2801 Founders Drive, Raleigh, NC 27695, USA*
[b] *Department of Economics, University of Connecticut, Oak Hall, 365 Fairfield Way, Storrs, CT 06269-1063, USA*
[c] *Department of Economics, Antai College of Economics and Management, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200030, China*

ARTICLE INFO

JEL classification:
Q51
Q58
R38

Keywords:
Agglomeration economies
Emission taxes
Emission standards
Firm relocation
Firm entry and exit
Effectiveness of environmental policy

ABSTRACT

This paper seeks to better understand the persistent environmental problems in urbanizing economies. We examine the effectiveness of environmental policy in an economy with agglomeration economies and endogenous firm relocation and entry/exit. We show that, although environmental regulation is effective in the short run, in the presence of agglomeration economies, spatial relocation of firms in response to environmental regulation can undermine the effectiveness of regulations, rendering them less effective or even ineffective. In fact, we show that regulation might even be counter-productive, i.e., exacerbate environmental problems, at certain stages of development. We present initial empirical evidence in the context of water pollution in China that demonstrates the importance of agglomeration economies in determining the impacts of environmental regulation.

1. Introduction

One of the most visible and highly publicized environmental problems in China, India and some other emerging economies is high concentrations of air pollution in their major cities. Almost all the top 25 most polluted cities in the world were located in those countries, with China and India accounting for most of them during the last ten years (IQAir 2022). In November 2021, schools in New Delhi were shut down for a week due to high levels of air pollution as the megacity was considering imposing a complete lockdown (Kim 2021). Water pollution is also a common problem, with India's Ganges River and China's Yangtze River often listed as the most polluted rivers in the world (Dhiraj 2017). In addition, environmental problems in those countries seem persistent, despite efforts to address them. For example, ever since Beijing was selected to host the 2008 Summer Olympic Games in 2001, the Chinese government has been trying to improve air quality in China. It has imposed many dramatic policy measures, including forcing heavily polluting

☆ An early version of this paper was presented as a keynote in the First Annual Conference of the Chinese Association of Environmental and Resource Economists, Beijing, China, August 24–26, 2017. We thank seminar participants at the University of Massachusetts, University of Kansas, Iowa State University, Renmin University of China, Xiamen University, Oregon State University, the 2019 AERE Summer Conference at Lake Tahoe, Nevada, and two anonymous referees and Co-Editor Martin Quaas for helpful comments and suggestions. Wang gratefully acknowledges financial support from the National Natural Science Foundation of China (Project #72273089). The authors declare no known interests related to this manuscript.

* Corresponding author.
E-mail address: junjie_wu@ncsu.edu (J. Wu).

https://doi.org/10.1016/j.jeem.2022.102754
Received 2 December 2021; Received in revised form 7 October 2022; Accepted 1 November 2022
Available online 5 November 2022
0095-0696/

Please cite this article as: JunJie Wu, *Journal of Environmental Economics and Management*, https://doi.org/10.1016/j.jeem.2022.102754

firms to shut down, imposing driving restrictions, and limiting new car purchases in some major cities (Zheng and Kahn 2013). Despite these efforts, air quality problems persisted in many Chinese cities. During this period total populations have more than doubled in many Chinese cities. The sheer increase in the volume of economic activity and the associated congestion and pollution have significantly impeded efforts to improve air quality in many Chinese cities.

In this paper we examine one factor that might contribute to why it can be so challenging to improve air quality in cities like Beijing and New Delhi during certain period of development, namely, the interplay between environmental regulation and agglomeration economies. We show that, in the presence of agglomeration economies, spatial relocation of firms in response to environmental regulation can undermine the effectiveness of regulations, rendering them less effective or even ineffective. In fact, we show that regulation might even be counter-productive, i.e., exacerbate environmental problems, at certain stages of development.

Broadly speaking, agglomeration economies are positive externalities generated when a large number of firms or individuals are located in close proximity to one another (*e.g.*, Henderson 2010). Agglomeration economies have long been a focal point of scholarly inquiry in regional and urban economics. The classic work has identified primary sources of agglomeration economies, including labor market pooling, input sharing, transport cost savings, knowledge spillovers, and consumer preferences for a variety of goods and services (see Duranton and Puga (2004) for a review of the literature). A large body of empirical literature has also tested the nature and scope of alternative sources of agglomeration economies and suggests that agglomeration economies are major determinants of the spatial distribution of economic activity.[1] When economic activity in turn generates pollution, these same agglomeration economies can have important environmental implications as well, particularly in rapidly urbanizing economies. For example, the resulting agglomeration may lead to both higher pollution in urban areas as well as more exposure as urban populations grow.

It is well recognized that agglomeration can lead to increased pollution and exposure (see literature review in Section 2). However, agglomeration can also affect the effectiveness of environmental regulation by affecting the cost of regulation and the benefit of agglomeration. For example, because firms in a larger market have easier access to information and technologies for pollution control, they may face lower compliance costs than firms in smaller markets, thereby increasing the benefits of agglomeration. In addition, environmental regulation aimed at reducing pollution can initially improve environmental quality in urban areas and thereby reduce the costs of agglomeration. In both cases, the effect of the regulation on the net benefits of agglomeration will in turn impact the concentration of economic activity and hence output and pollution. This indirect feedback effect through agglomeration can offset the direct effect of the regulation and thus reduce its effectiveness. However, to the best of our knowledge, there has been little, if any, research that has analyzed the effect of agglomeration economies on the effectiveness of environmental policy through these linkages.

A primary objective of this paper is to understand how endogenous responses to agglomeration economies can affect the effectiveness of environmental policy. To address this question, we first develop a theoretical model of an economy with pollution and agglomeration economies and use the model to analyze the effectiveness of environmental policy. We consider two types of regulatory policies: a uniform standard that is applied in all regions, and a differentiated standard that is higher in more polluted regions. For each, we contrast the effectiveness of policy under the typical short run assumption of fixed firm location with its effectiveness if entrepreneurs facing agglomeration economies can endogenously relocate their firms within the country in response to changes in environmental quality (as well as prices and profits) in those regions. The importance of environmental quality as a location factor has been noted by, for example, Burkitt and Spegele (2013), who report that having to breathe the severely polluted air in cities such as Beijing and Shanghai has led some entrepreneurs to favor locating in other less polluted cities. We also consider policy effectiveness with not only endogenous relocation but also endogenous entry/exit. This reflects the fact that emerging economies are often characterized by not only rapid urbanization but also an increasing number of new firms. For example, China's "economic miracle", a shift from a largely agrarian society to an industrial powerhouse, was started by a fundamental reform in its economic system in 1978. This reform removed barriers for laborers to establish private enterprises and led to a dramatic increase in the number of firms in China.

The theoretical model suggests that consideration of agglomeration economies and firm relocations and entry/exit can change some of the classic results regarding the impact of environmental policy. In particular, we find that, although standards commonly used for emission control are effective in reducing aggregate pollution in the short run when both the number and location of firms are fixed, they may not have the intended result once endogenous relocation and entry/exit are considered. In particular, with endogenous relocation a regulatory standard will generally be less effective and might actually exacerbate pollution problems (*i.e.*, lead to more total emissions, higher spatial concentration of pollution, and/or larger pollution damages). The key reason is that adoption of cleaner technology makes agglomeration less costly, which can lead to increased spatial concentrations of firms and more economic activity that can offset the direct environmental benefit of regulation.[2] As a result, an "environmental stagnation" or "environmental trap" can occur in economies that are partially agglomerated, a situation characteristic of rapidly urbanizing countries. If over time firms also make endogenous entry/exit decisions, standards can still be counterproductive depending on how they affect firms' costs.

We also take an initial step to explore the interaction between environmental regulation and agglomeration economies empirically,

[1] Rosenthal and Strange (2004) provide an extensive review of empirical studies on the scope and nature of agglomeration economies as of the early 2000s. Since then many new studies have been conducted in this area, including Moretti (2004), Shapiro (2006), Glaeser and Gottlieb (2008), and Ellison et al. (2010). These more recent studies typically use unique data and innovative strategies to more accurately measure some of the specific sources of agglomeration economies.

[2] This feedback effect is akin to the "rebound effect" of energy efficiency standards, whereby improvements in the energy efficiency of appliances or cars induces greater use that can offset, at least partially, the reductions in energy use that the higher standards might otherwise induce. We thank an anonymous referee for noting this analogy. For a review of the literature on the rebound effect for energy efficiency standards, see Gillingham et al. (2016).

in the context of water pollution regulation in China. We first provide evidence consistent with our theoretical result that the impact of environmental regulation in a region depends on its agglomeration economies. We show that agglomeration economies reduce the impact of regulations on the number of firms in a county and work to undermine regulation's impact on pollution reduction. In our theoretical model this is due, in part, to the impact of agglomeration on the cost of reducing emissions. We therefore also present empirical evidence supporting this assumption. In particular, we show that firms that discharge Chemical Oxygen Demand (COD) pollution face lower abatement costs when located in counties with larger agglomeration economies. Both the theory and the empirical analysis highlight the importance of incorporating agglomeration economies when considering the impact of environmental policy and provide a possible explanation for, or at least a factor contributing to, the challenges for environmental improvements in rapidly urbanizing economies.

The paper is organized as follows. In Section 2 we provide a brief overview of the related literature. Section 3 presents the basic setup of the model, and Section 4 characterizes the equilibrium outcomes in the short run as well as in the intermediate and the long run without environmental regulation, which serves as a baseline for evaluating the effect of environmental policy. Section 5 analyzes the effectiveness of uniform environmental policy, while Section 6 focuses on the effect of a spatially differentiated policy. Section 7 presents empirical evidence that supports the importance of the link between environmental regulation and agglomeration in the context of water pollution control in China. Section 8 summarizes our results and conclusions.

2. Related literature

A large literature within urban economics has examined both the centripetal forces for agglomeration and the centrifugal forces for decentralization, as well as the effect of these forces on the spatial distribution of economic activity and economic growth (Duranton and Puga 2004; Wu 2019). Glaeser (1998, p. 140) notes that the centripetal forces favoring agglomeration stem from reductions in the "transport costs for goods, people and ideas", while centrifugal forces favoring decentralization include "health costs, pollution, congestion, crime and social problems." Previous studies that have modeled pollution as a centrifugal force that causes relocation away from more polluted areas include Henderson (1977), Arnott, et al. (2008), Kyriakopoulou and Xepapadeas (2017), Regnier and Legras (2018), Borck and Tabuchi (2019), and Pflüger (2021). These papers generally focus on land market equilibria when workers can live in the periphery of a city and commute to work in a central business district, where the equilibria are driven by (among other things) a tradeoff between commuting costs and pollution.

In contrast to the above models that focus on commuting, models from the "new" economic geography are more general equilibrium and focus on the role of immobile factors and market size, as well as economies of scale and imperfect (monopolistic) competition (*e.g.*, Krugman 1998). A small number of studies have incorporated pollution and agglomeration into this modeling framework.[3] For example, van Marrewijk (2005) uses a two-factor/two-sector economic geography model to show that agglomeration is less attractive and hence less likely when the sectors generate pollution. Hosoe and Naito (2006), Lange and Quaas (2007) and Wu and Reimer (2016) use a similar model to characterize different equilibria of firm concentrations and show that pollution affects the extent of agglomeration that occurs in equilibrium. However, these studies focus on the impacts of pollution on the spatial distribution of economic activity, rather than on the effectiveness of environmental policy, which is the focus of this paper.

There is also a large literature within environmental economics that has examined the effect of environmental policy on various economic activities, including firm relocation (Markusen et al., 1993; List et al. 2003), firm births/deaths (Becker and Henderson 2000; List et al. 2004), competitiveness (Dechezlepretre and Sato 2017), capital stock (Greenstone 2002), employment growth (Walker 2011; Kahn and Mansur 2013), foreign direct investments (Cai et al., 2016), and exports (Hering and Poncet 2014). Some of these studies consider spatial dimensions of polluting activities, focusing on the spatial variation in marginal pollution damage or marginal pollution control cost and how consideration of the spatial variation could improve the efficiency of the policy (Mendelsohn 1986; Muller and Mendelsohn 2009; Cullen 2013; Holland and Yates 2015). However, they do not generally focus on the role of agglomeration. For example, in an important study of a low carbon fuel standard (LCFS), Holland et al. (2009) show that limiting producers' emissions per unit of output can raise total carbon emissions, rather than reduce them, and as a result an emission standard can be counterproductive (a result that is similar to ours). However, agglomeration is not the driver of their results. Policy-oriented studies that include both pollution and agglomeration typically focus on the design of optimal policies that include environmental and/or land (or factor) use policies aimed at inducing optimal land use or city size (*e.g.*, Henderson 1977; Arnott et al., 2008; Kyriakopoulou and Xepapadeas 2017; Regnier and Legras 2018; Pflüger 2021). In contrast, we are primarily interested in the effectiveness of environmental regulation, which has been an issue in many developing countries like China and India.

Finally, several studies have specifically focused on challenges for environmental management in developing countries, where "informal regulations" often exist (Pargal and Wheeler 1996). Some of these studies examine self-regulation (Christmann and Taylor

[3] In related work, Elbers and Withagen (2004) consider the pollution haven hypothesis within the context of "new trade theory" based on a spatial model of monopolistic competition and show that pollution can countervail clustering that would otherwise occur. Zeng and Zhao (2009) similarly show that agglomeration forces can alleviate the pollution-haven effect. For a review of the pollution haven literature, see Levinson and Taylor (2008). The pollution haven phenomenon is a special case of "emission leakage" in which firms relocate in response to spatially differentiated policies. Leakage can substantially offset, or even reverse, the reductions in emissions achieved in the regulated sector (Holland 2012; Fowlie and Muller, 2019; Fowlie et al., 2016). Another related literature examines the effect of pollution on trade. For a recent review of literature on trade and the environment, see Chernwichan et al. (2017). However, neither the leakage literature nor the trade-environment literature focuses on the role of agglomeration.

2001), and others focus on the impact of regulation, for example on foreign direct investment (Cai et al., 2016).[4] Again, however, this work does not explicitly consider the role of agglomeration economies.

3. Model setup

Consider an economy with two physically identical[5] regions, East (E) and West (W), and two sectors, manufacturing and agriculture. The manufacturing sector produces differentiated goods, and the agricultural sector produces a homogenous, numéraire good. There are two types of individual consumers in the economy: laborers and entrepreneurs. Each entrepreneur owns a single firm that produces a unique manufactured good (*i.e.*, a unique variety) in a single plant located in their region. We assume entrepreneurs are mobile between the regions, but they live in the region where their firms are located. This assumption likely holds for a vast majority of firms in developing countries, which tends to be small family-owned firms. For example, in China estimates suggest that a 1% increase in the total number of firms is associated with a 0.92% increase in the total number of firm owners.[6] In contrast to entrepreneurs, laborers are immobile and evenly distributed across the two regions. Let L be the total mass of laborers and N be the total mass of entrepreneurs in the economy. Then $M = L + N$ is the total (exogenous) mass of population (consumers) in the economy. We take L and N as fixed in the short run, but allow for the possibility that over time laborers can choose to become entrepreneurs, thereby increasing N and decreasing L in the long run. Let λ denote the share of entrepreneurs located/living in the East. Then the total masses of population in the East and West are given by $M_E = \lambda N + 0.5(M - N)$ and $M_W = (1 - \lambda)N + 0.5(M - N)$, respectively. Note that $\lambda = 0.5$ corresponds to a totally dispersed economy, while $\lambda = 0$ or $\lambda = 1$ implies total agglomeration of manufacturing into one of the two regions. Without loss of generality, we assume that any agglomeration that occurs happens in the East. Thus, we assume throughout that $\lambda \geq 0.5$.

Laborers inelastically supply labor to the agricultural sector in both regions in exchange for wages. The agricultural sector is perfectly competitive in both regions and has a constant return to scale technology that requires one unit of labor (*i.e.*, one laborer) to produce one unit of agricultural output. This, coupled with the choice of the agricultural good as the numéraire, implies that in equilibrium the wage paid to each laborer is equal to one in both regions.

The production process is identical for all manufacturing firms and is represented by a cost function that includes variable and fixed costs. The variable cost reflects the cost of raw materials used in manufacturing. We assume production of each unit of manufacturing good requires a given amount of raw materials (in addition to the skill and time of the entrepreneur) that can be purchased at a constant price. This implies that the marginal cost of production is constant. The presence of a fixed cost, reflecting the needed capital investment and overhead cost, implies that the technology exhibits economies of scale.

We capture agglomeration economies in two ways. First, we assume that a firm can sell its product to consumers in either region, but it is costly to transport goods from one region to the other. Transportation costs provide an incentive for firms to locate where demand is relatively larger, thereby providing a potential rationale for agglomeration (Krugman 1991). Second, we assume the fixed cost of production decreases as the total number of firms increases in a region.[7] This reflects the idea that firms can benefit from co-locating near other firms, due, for example to sharing fixed inputs, knowledge spillovers, or savings in the cost of supporting services. Specifically, we assume a firm's costs in region r equal $C_r(q) = wq + K(1 - \varphi N_r)$, where q is the firm's output, w is the marginal cost of production, N_r is the total number of firms in region r, and $K > 0$ and $\varphi \geq 0$ reflect scale economies and co-location benefits, respectively, with φ defined such that $\varphi N_r < 1$ for any $N_r \leq N$.[8]

Production of manufactured goods generates a fixed amount of emissions of a given pollutant per unit of output, given by z, which is identical for all varieties of the manufactured good (*i.e.*, all varieties are equally polluting). Thus, z represents emissions intensity. Below we focus on regulations that reduce emissions intensity, which can vary significantly across developed and developing countries. For example, in 2014 the CO_2 emission intensity was nearly seven times higher in China than in the U.S. (0.178 kg per million dollars of GDP (constant 2010 US dollars) in the U.S. compared to 1.235 in China) (World Bank, 2018). Changes in emissions intensity can also explain overall environmental improvements. For example, Shapiro and Walker (2018) analyze the changes in air pollution

[4] For a recent survey of the empirical literature on the effectiveness of specific regulatory actions in developing countries, see Blackman et al. (2018).

[5] This assumption implies that neither region has a natural comparative advantage. For a model where natural comparative advantage can drive the spatial dispersion of economic activity, see Kyriakopoulou and Xepapadeas (2013).

[6] This estimate is based on simple regressions we ran using city-level data for China. The city-level measures were calculated using three widely-used datasets: China Population Census, China City Statistical Yearbook, and the China Economic Census 2004. First, we used the 20% sample of the 2005 1% Mini Population Census to calculate the share of individuals that were identified as "employer" in each of the 277 prefecture-level cities. The total number of employers (firm owners) was then obtained by multiplying this estimated share by city population from the China City Statistical Yearbooks. Data on the number of firms for each city were obtained from the China Economic Census 2004. Note that due to data availability, the regression analysis was conducted using data from two different years (2004 and 2005).

[7] We thus take a dual approach to representing agglomeration economies (by including them in the firm's cost function). Some studies take a primal approach and include agglomeration economies in the firm's production function (*e.g.*, Fugita and Ogawa, 1982; Lucas and Rossi-Hansberg 2002; Wang and Wu 2011). The dual approach embodies agglomeration economies that affect production, coupled with cost-minimizing behavior. Both approaches lead to an indirect profit function that shows that profit increases with agglomeration.

[8] An alternative specification would have agglomeration economies impacting variable rather than fixed costs. This would be appropriate, for example, for industries where agglomeration increases worker productivity (Duranton and Puga 2004). As we show in the appendix, our basic results hold in this case as well.

emissions from U.S. manufacturing from 1990 to 2008 and find that the large reductions during this period were primarily driven by reductions in emissions intensity rather than changes in output or in the composition of products produced. These reductions in emissions intensity could be driven by technological progress or adoption of new technologies, which could in turn affect fixed costs. Thus, although we initially treat K as a parameter, in our analysis of regulation below, we allow K to vary with firms' emissions intensity.

A key aspect of emissions is that different types of pollutants have different levels of mobility. For example, greenhouse gas emissions may be more easily dispersed across regions than local water pollution or waste disposal. To capture this, we assume that, of the z units of the pollutant emitted per unit of manufacturing output, a fraction γ stays in the home region where the firm is located, and the rest drifts to the other region, where $0.5 \leq \gamma \leq 1$. Thus, γ measures the stationarity of emissions, with $\gamma = 1$ indicating emissions remain entirely local and $\gamma = 0.5$ indicating that the emissions are dispersed evenly across both regions, which implies that the location of the production activity generating those emissions is unimportant.[9]

All consumers are assumed to have identical preferences defined by a quadratic utility function[10]

$$U = \alpha \int_0^N q(i)di - \frac{\beta}{2} \int_0^N q^2(i)di + q_A - \left(Z + \delta Z^2\right), \tag{1}$$

where α and β are positive parameters that measure the intensity of consumer preferences for manufactured goods (relative to the agricultural good) and variety, respectively; $q(i)$ is the consumption of the i-th variety of the manufacturing good; q_A is the consumption of the agricultural good; Z is the level of pollution in the region where the consumer lives (which could differ from emissions generated in that region if pollution is mobile); and $D(Z) = Z + \delta Z^2$ is the pollution damage function, with $\delta \geq 0$ implying that the marginal pollution damage is non-decreasing.

Consumers choose the consumption level of each good to maximize their utility subject to a budget constraint, taking the level of pollution as given. This yields the following individual demand functions for each of the manufacturing goods and the corresponding indirect utility function:

$$q(i) = \frac{\alpha - p(i)}{\beta} \tag{2}$$

$$V = \frac{N\alpha^2}{2\beta} - \frac{\alpha}{\beta}\left(\int_0^N p(i)di\right) + \frac{1}{2\beta}\left(\int_0^N p^2(i)di\right) + Y + Y_0 - \left(Z + \delta Z^2\right), \tag{3}$$

where $p(i)$ is the price of the i-th variety of the manufactured good; Y is the consumer's income (from wages for laborers and from profits for entrepreneurs); and Y_0 is the individual's initial endowment of the numéraire good, which is assumed to be large enough to ensure some consumption of that good.

The manufacturing market is characterized by monopolistic competition. Each firm faces a downward-sloping demand curve given by (2) and chooses its price to maximize its total profit. The number of firms is large enough that each firm can ignore its influence on, and reactions from, other firms. Thus, each firm in region r solves the following problem:

$$Max_{p_{rr}, p_{ro}}\ \pi_r \equiv (p_{rr} - w)q_{rr}M_r + (p_{ro} - w - \tau)q_{ro}M_o - K_r,$$

where $r, o = E, W$ and $r \neq o$; p_{rr} and p_{ro} are the prices of a manufactured good produced in region r sold in the home region (r) and the other region (o), respectively; K_r is the fixed cost in region r; and $\tau \geq 0$ is the cost of transporting one unit of any variety from one region to the other. First-order conditions of the firm's maximization problem result in the following (identical) pricing strategies for each firm:

$$p_{rr} = \frac{\alpha + w}{2} \quad \text{and} \quad p_{ro} = \frac{\alpha + w + \tau}{2}. \tag{4}$$

A firm's markup over the variable and transport cost for goods sold in the other region is positive if and only if $\tau < (\alpha - w)$. Thus, throughout the analysis below we assume $\tau < (\alpha - w)$ to ensure inter-region trade.

4. Equilibrium outcomes without environmental regulation

As a benchmark or baseline, we first derive the economic and environmental outcomes without any policy, in the short run (where location is fixed), the intermediate run (where firms endogenously choose locations), and in the long run (where firm entry/exit can occur, *i.e.*, laborers can become entrepreneurs).

[9] Kyriakopoulou and Xepapadeas (2011, 2013) specify how emissions diffuse over space in a landscape. For simplicity, we do not model diffusion within a given region because our regions are dimensionless. In addition, to simplify the analysis, we assume the air pollutant does not accumulate in the air and does not move primarily to the other region (which would imply $\gamma < 0.5$).

[10] We adopt a simplified version of the quadratic utility function of Ottaviano et al. (2002) to simplify the model and notation. The analysis can be carried out in a similar fashion with the more general specification without changing the fundamental results of this paper.

4.1. Short run equilibrium

Given the demand functions in (2) and the prices in (4), the total production of manufactured goods in region r, Q_r, is

$$Q_r = N_r(q_{rr}M_r + q_{ro}M_o) = \frac{N_r}{2\beta}[\tau(N_r - 0.5N) + M(\alpha - w - 0.5\tau)]. \tag{5}$$

(All derivations, as well as the proofs of all propositions, are given in the appendix, which is available online.) Substituting (2) and (4) into profit gives the income for an entrepreneur living and producing in region r:

$$Y_r = M\overline{\pi} + \Delta\pi(N_r - 0.5N) - K(1 - \varphi N_r), \tag{6}$$

where $\Delta\pi = \tau(\alpha - w - 0.5\tau)/2\beta$ and $\overline{\pi} = [(\alpha - w)^2 + (\alpha - w - \tau)^2]/2\beta$. The difference in income for entrepreneurs in the two regions is

$$(Y_E - Y_W) = 2(\Delta\pi + \varphi K)N(\lambda - 0.5). \tag{7}$$

In the presence of agglomeration economies, *i.e.*, when $\tau > 0$ (and hence $\Delta\pi > 0$) or $\varphi > 0$, $(Y_E - Y_W)$ is positive when $\lambda > 0.5$, *i.e.*, an entrepreneur will earn more profit when located in the larger market (the East). The larger the transportations costs τ or co-location benefits φ, the larger the income difference. The aggregate income for entrepreneurs in the economy is

$$Y_T = \lambda NY_E + (1-\lambda)NY_W = N[M\overline{\pi} - K(1 - 0.5\varphi N)] + 2(\Delta\pi + \varphi K)N^2(\lambda - 0.5)^2, \tag{8}$$

which in the presence of agglomeration economies is increasing in λ when $\lambda > 0.5$ (*i.e.*, when firms are more concentrated in one region).

Production decisions in the two regions, combined with emissions per unit of output and the mobility of emissions, determine the level of pollution in each region, which is given by:

$$Z_r = \gamma zQ_r + (1-\gamma)zQ_o, \quad r, o = E, W. \tag{9}$$

These pollution levels in turn determine aggregate pollution, the difference in pollution levels (a proxy for spatial concentration of pollution), and total pollution damages, as follows:

$$Z_E + Z_W = \frac{z}{2\beta}\left[2\tau N^2(\lambda - 0.5)^2 + MN(\alpha - w - 0.5\tau)\right], \tag{10}$$

$$Z_E - Z_W = \frac{(2\gamma - 1)z}{\beta}[2M(\alpha - w) - (M - N)\tau]N(\lambda - 0.5), \tag{11}$$

$$TD = \left(Z_E + \delta Z_E^2\right)M_E + \left(Z_W + \delta Z_W^2\right)M_W = 0.5M(Z_E + Z_W) + (Z_E - Z_W)N(\lambda - 0.5)$$

$$+\delta(Z_E + Z_W)(Z_E - Z_W)N(\lambda - 0.5) + 0.25M\delta\left[(Z_E + Z_W)^2 + (Z_E - Z_W)^2\right]. \tag{12}$$

Equation (10) implies that, in the presence of transportation costs, the total amount of pollution increases as firms become more concentrated. Intuitively, given the number of firms, the overall price reaches the lowest point when all firms locate in one region due to the saving of transport costs. As a result, the total demand and hence aggregate production and pollution are maximized with full agglomeration. Moreover, if $\gamma > 0.5$, the spatial concentration of pollution and total pollution damages are increasing with agglomeration of firms in one region. These results also imply that aggregate pollution, the spatial concentration of pollution, and aggregate damages from firms' pollution are all minimized when firms are perfectly dispersed, i.e., when $\lambda = 0.5$. Thus, from a purely environmental perspective, perfect dispersion of economic activity is optimal.[11] This is true regardless of the extent to which pollution is local vs. regionally dispersed, i.e., regardless of the magnitude of γ, and even when the damage function is linear (i.e., $\delta = 0$).

4.2. Intermediate run equilibrium

The above analysis assumes a given distribution of firms and shows how both economic and environmental outcomes depend on

[11] It is important to note that here we focus on pollution from production. In particular, we do not consider emissions from housing, commuting, or the transportation of goods. When these consumption-based pollution sources are considered, the spatial distribution of people that minimizes overall pollution may be different (Borck and Pflüger 2019). Evidence of spatially-driven consumption-based externalities certainly exists. For example, Glaeser and Kahn (2010) analyze the correlation between development density and gasoline usage and find that the average household living in a census tract with fewer than 1000 people per square mile uses 70% more gasoline annually than the average household living in a census tract with more than 10,000 people. Similarly, Grazi et al. (2008) and Vance and Hedel (2008) find that households living in higher density places generate lower greenhouse gas (GHG) emissions. More generally, however, the literature on whether per capita emissions or ambient pollution levels increase or decrease with a city's population or density is mixed (see Ahlfeldt and Pietrostefani 2019; Borck and Pflüger 2019). For example, Carozzi and Roth (2019) find a positive relationship between city density and ambient particulate pollution. How pollution varies with city size can have implications for the optimal city size (*e.g.*, Borck and Tabuchi 2019).

that distribution. In general, however, if entrepreneurs are mobile, they will be able to choose their locations. In this subsection, we consider how agglomeration economies, along with pollution and its characteristics, affect the equilibrium distribution of firms through their impacts on the location decisions of entrepreneurs, and how this, in turn, affects environmental outcomes.

To capture endogenous location decisions, we assume that entrepreneurs choose to locate in the region that gives them the highest utility, taking into consideration the profit they will earn from production, the price index of consumption goods produced in both regions, and environmental quality in the two regions. Equation (3) implies that the utility difference for entrepreneurs located in the two regions depends on the number and distribution of firms and is given by:

$$\Delta V \equiv V_E - V_W = -C\left[(2\gamma-1)\delta\tau z^2 N^2(\lambda-0.5)^2-\theta\right]N(\lambda-0.5), \tag{13}$$

where

$$C=\frac{1}{\beta^2}\left[M(\alpha-w)-0.5(M-N)\tau\right]>0 \tag{14}$$

$$\theta=\frac{2}{C}(3\Delta\pi+2\varphi K)-(2\gamma-1)z[\beta+0.5z\delta MN(\alpha-w-0.5\tau)]. \tag{15}$$

A distribution of firms between the two regions is in equilibrium if no entrepreneur has an incentive to move to the other region, which can occur only if utility is the same in the two regions or is at least as high in one region when all entrepreneurs locate in that region. In addition, an equilibrium is said to be stable if for a small deviation from the equilibrium, entrepreneurs would have an incentive to move in a way that would bring the spatial distribution of firms back to the equilibrium.[12]

In the absence of pollution, *i.e.,* when $z=0$, the model gives the standard results from the literature (*e.g.*, Ottaviano et al., 2002), namely, the region with the larger initial share of the manufactured sector attracts the whole sector because of agglomeration economies, leading to full agglomeration being the only stable equilibrium. However, here we are interested in characterizing the equilibrium when $z>0$, which we assume throughout the remainder of the paper. In addition, throughout this and the following section we assume that the damage function is strictly convex ($\delta>0$) and transportation costs are positive ($\tau>0$).[13]

We begin by considering the case where pollution is perfectly dispersed, *i.e.*, $\gamma=0.5$, where we can show the following[14]

Proposition 1. *If $\gamma=0.5$, then the stable equilibrium is full agglomeration, i.e., $\lambda^*=1$.*

When $\gamma=0.5$, pollution is a "global" public bad, implying the location of emissions does not matter and therefore the distribution of firms has no effect on damages in the two regions. (An example is carbon dioxide emissions.) In this case, firms simply locate to take advantage of agglomeration economies and the equilibrium is identical to the case without pollution.

We turn now to characterizing the equilibrium for situations where pollution is not perfectly dispersed, which is our primary interest. In this case, assuming again (without loss of generality) that any agglomeration that occurs is in the East, we have the following:

Proposition 2. *If $\gamma>0.5$, the equilibrium distribution of firms in the intermediate run is given by*

$$\lambda^*=\begin{cases}1 & \text{if } z\le\bar{z}\\ 0.5+\sqrt{\widetilde{\theta}} & \text{if } \bar{z}<z<z\\ 0.5 & \text{if } z\ge\bar{\bar{z}}\end{cases}\;. \tag{16}$$

where $\widetilde{\theta}=\frac{\theta}{(2\gamma-1)\delta\tau z^2N^2}$, and $\bar{z}$ and z solve $\widetilde{\theta}=0.25$ and $\widetilde{\theta}=0$, respectively.[15]

Proposition 2 implies that the type of equilibrium that emerges – full dispersion, partial agglomeration, or full agglomeration – depends not only on agglomeration economies but also on pollution and its characteristics (i.e., on the emissions intensity (z), the mobility of emissions (γ), and the convexity of marginal pollution damage (δ)). Because of our interest in the role of pollution and regulations that can change pollution intensity, Proposition 2 focuses on the importance of z. (The boundaries between the equilibria can also be expressed in terms of other parameters.) In particular, it shows that, when $\gamma>0.5$, the relative value of z determines the distribution of the firms between the two regions. When emissions intensity is sufficiently low, the equilibrium outcome is full agglomeration (because of agglomeration economies), which is the standard result from the urban economics literature. However, as z increases, the equilibrium distribution of firms moves from full agglomeration to partial agglomeration/dispersion and ultimately to full dispersion. Intuitively, when emissions cause more damage locally (i.e., $\gamma>0.5$), there is a tradeoff between agglomeration benefits and pollution costs. When emission intensity is relatively low, agglomeration benefits will dominate and agglomeration will

[12] We will define dynamic equilibria and stability explicitly when considering entry/exit/relocation decisions in Section 4.3.

[13] If $\delta=0$ or $\tau=0$, then the possible equilibria are either full agglomeration, full dispersion, or indeterminate (any value of λ), depending on whether θ is positive, negative, or zero.

[14] Note that here and in Proposition 2 below $\lambda^*=0$ would also be an equilibrium because the two regions are symmetric. This equilibrium is ruled out by our assumption that the East is larger whenever the two regions have different populations.

[15] Below, when we allow fixed costs to vary with z, the equations defining $\bar{z}$ and $\bar{\bar{z}}$ become $\widetilde{\theta}(z,\varphi,K(z),N)=0.25$ and $\widetilde{\theta}(z,\varphi,K(z),N)=0$, respectively. A similar comment applies to the expressions in Proposition 3. See proofs in the Appendix for details.

occur. However, when emission intensity increases, pollution will act as a centrifugal force, initially leading to a partial agglomeration equilibrium that balances these two forces. When emissions intensity is sufficiently high, environmental costs will outweigh agglomeration benefits, and firms will be perfectly dispersed between the two regions. Thus, when emissions are not perfectly dispersed, the strength of the pollution effect, i.e., the magnitude of z, will determine the extent of agglomeration. This result is consistent with Lange and Quaas (2007), who show that these same three possible equilibria can occur, depending on the damages from local pollution.

We are particularly interested in the range of partial agglomeration, since this is likely to characterize many developing countries.[16] Proposition 2 implies that under partial agglomeration, we have

$$\frac{\partial \lambda^*}{\partial \varphi} > 0, \left.\frac{\partial \lambda^*}{\partial K}\right|_z > 0 \tag{17}$$

which imply that larger co-location benefits and scale economies will lead to greater concentration of firms.[17] Therefore, although agglomeration economies from co-location benefits and scale economies do not directly affect environmental outcomes (see (10)-(12)), they have an indirect environmental impact through their impact on the location of firms. In particular, through this indirect effect, an increase in agglomeration economies can lead to more aggregate pollution, higher spatial concentration of pollution, and larger pollution damage.

Agglomeration is also affected by the different characteristics of pollution. Again by Proposition 2, under partial agglomeration we have that

$$\left.\frac{\partial \lambda^*}{\partial z}\right|_K < 0, \frac{\partial \lambda^*}{\partial \gamma} < 0, \frac{\partial \lambda^*}{\partial \delta} < 0. \tag{18}$$

These results suggest that, ceteris paribus, higher emission intensity, lower emission mobility, and greater convexity of marginal pollution damage all lead to less concentration of firms in one region. Intuitively, as the emission intensity z increases, benefits from locating in the East (the larger region) decrease because pollution is more concentrated in the region (i.e., ceteris paribus, $(Z_E - Z_W)$ increases with z). This will in turn increase the number of firms located in the West. In addition, the net benefit of agglomeration decreases as emissions become more stationary (i.e., γ is larger) or when the marginal pollution damage function is more convex (i.e., δ is larger).[18] Therefore, an increase in γ or δ will lead to a reduction in λ^*. These results suggest that pollution intensity and characteristics also affect total pollution, the spatial concentration of pollution, and total pollution damages indirectly through changes in λ^* (see equations (10)-(12)).

4.3. Long run equilibrium

In this subsection we explore how firm entry/exit would affect the equilibrium outcomes in the long run. To model firm entry/exit, we assume individuals can choose between being a laborer and being an entrepreneur, but being an entrepreneur involves significant transaction costs, denoted by μ, which could represent costs associated with, for example, migration, education and training, stress and additional time/effort, or different lifestyle choices. Thus, a laborer will want to be an entrepreneur if and only if the additional income that can be earned by becoming an entrepreneur is sufficiently high to offset these costs.[19] Since profits are higher in the East, this additional income is given by $Y_E - 1$, which according to (6) depends on both the current number of firms in the industry, N, as well as the distribution of firms, λ. Thus, a laborer will be better off as an entrepreneur if and only if $\mu < Y_E - 1$, or, equivalently, if $Y^* \equiv (\mu + 1) < Y_E$. We assume Y^*, or equivalently μ, varies across the population, with a cumulative distribution function $H(Y^*)$ and a probability distribution function $h(Y^*)$ over the interval $[1, \overline{Y}]$ (where the lower end of the interval corresponds to $\mu = 0$). Thus, at a given time, the total number of people in the population who would be better off as an entrepreneur equals

[16] There is strong empirical evidence that economies move through a stage of partial agglomeration as they develop. For example, Latzko (2018) estimates the Gini coefficients for manufacturing output for the period 1820 to 2010 and finds that the geographic concentration of manufacturing output increased rapidly as the real per-capita GDP increased from \$2000 to \$5000, flattened out between \$5000 and \$10,000, and then began to fall. Similarly, Kim (1995) develops a concentration index to measure the geographic concentration of economic activity and finds an inverted U-shaped relationship between real per-capita GDP and the geographic concentration of economic activity. Finally, using the Ellison-Glaeser index to measure the geographic concentration of economic activity, Wu et al. (2022) find a negative correlation between the concentration of economic activity and real per-capita GDP for 1972–1992, suggesting that higher levels of income are accompanied by more spatial dispersion.

[17] As noted, below we let K change when z changes. Here, however, the second partial in (17) should be interpreted as the effect of an increase in fixed costs (and hence scale economies) that is unrelated to any change in z. Similarly, the first partial in (18) should be interpreted as the direct effect of an increase in z, unrelated to any associated increase in fixed costs.

[18] Note that, when damages are strictly convex, starting from an equal distribution of pollution, any shifting of pollution from one region to the other will increase damages more in the receiving region than it will decrease damages in the other region.

[19] Our approach to modeling firm entry and exit differs from Melitz (2003), who emphasizes endogenous selection of productivity heterogeneity among firms. In this paper we focus on heterogeneity in laborers' reservation income for them to become entrepreneurs in an emerging economy, which is typically characterized by a large number of new entrepreneurs.

$$\widetilde{N} = \int_1^{Y_E} dH(Y^*) = H(Y_E(\lambda, N)). \tag{19}$$

If the existing number of entrepreneurs N is less than $\widetilde{N}$, some laborers would be better off becoming entrepreneurs and we would expect entry to occur. On the other hand, if N is greater than $\widetilde{N}$, some entrepreneurs would be better off as laborers, thereby inducing exit. The difference between N and $\widetilde{N}$ thus drives the dynamics of firm entry and exit. The following equation of motion can be used to capture changes over time in the total number of entrepreneurs in the economy:

$$\frac{dN(t)}{dt} = G(\widetilde{N} - N), \tag{20}$$

where $G'(\bullet) > 0, G(0) = 0$, and $G'(0) = 1$. This assumes that as laborers enter or entrepreneurs exit in response to the difference, the incentives to enter or exit decrease, and that the larger the difference between N and $\widetilde{N}$, the faster the total number of firms changes.

To characterize the dynamics of relocation, we follow a well-established tradition in the migration literature by assuming that incentives to move from one location to another increase with differences in utility across the regions (Tabuchi and Thisse 2006). This implies the following equation of motion for firm relocations:

$$\frac{d\lambda(t)}{dt} = F(\lambda(1-\lambda)\Delta V(\lambda, N)), (21)$$

where $F(0) = 0, F'(\bullet) > 0$, and $F'(0) = 1$.

The dynamic system (20) and (21) determines the evolution of N and λ over time, which in turn drives the trajectories of economic and environmental outcomes. Under partial equilibrium, the long run equilibrium requires that entrepreneurs have the same utility in the East as in the West (i.e., $\Delta V = 0$) and that each individual is in the occupation (laborer or entrepreneur) that gives that person the higher utility, given their value of μ (i.e., $N = \widetilde{N}$). As in the intermediate run, we can show that, when pollution is not perfectly dispersed, the long run steady state equilibrium, defined by $\frac{dN(t)}{dt} = 0$ and $\frac{d\lambda(t)}{dt} = 0$, depends on the pollution intensity z, where a sufficiently low level of z leads to full agglomeration, a sufficiently high level of z leads to complete dispersion, and intermediate levels of z lead to partial agglomeration. In particular, we can show the following:

Proposition 3. *. Suppose $\gamma > 0.5$. There exist $\overline{z}'$ and z' such that, when $\overline{z}' < z < z'$, the solution to the following two equations:*

$$\lambda = 0.5 + \sqrt{\widetilde{\theta}}$$

$$N = H(Y_E(\lambda, N))$$

defines a unique long run stable steady state equilibrium (λ^*, N^*) with partial agglomeration, i.e., $\lambda^* < 1$.

The intuition behind this result is similar to that behind Proposition 2. Note that both λ^* and N^* are functions of z.

5. Effectiveness of uniform environmental policy

We turn now to the impact of environmental regulation on the equilibrium outcomes. We consider first a regulation that requires all manufacturing firms in both regions to adopt a cleaner production technology or, equivalently in terms of our model, sets an emissions standard that limits allowable emissions per unit of output, z. (Qualitatively similar conclusions hold for an emissions tax, as shown in the appendix.) We assume throughout that the policy is binding and hence we can treat the policy as requiring a reduction in z. In addition, we assume that compliance with the standard requires the purchase and installation of new equipment that results in an increase in firms' fixed costs. We capture the increase in fixed costs by now assuming that the fixed cost (FC) depends on z, *i.e.*, $FC_r = K(z)(1 - \varphi N_r)$ for $r = E, W$, where $K'(z) < 0$.[20]

5.1. Uniform policy in the short run

The short-run environmental effect of the emission standard can be derived directly from equations (10)-(12). Holding the total number of firms (N) and firm locations (λ) constant, we can easily derive the following:

$$\frac{\partial(Z_E^* + Z_W^*)}{\partial z} > 0, \frac{\partial(Z_E^* - Z_W^*)}{\partial z} > 0, \frac{\partial TD^*}{\partial z} > 0. \tag{22}$$

These results suggest that an emission standard that requires a reduction in emission intensity z will have the expected (intended)

[20] Note that, if environmental regulation promotes innovation that offsets the cost of compliance (Porter 1990), then it is possible that the standard will decrease firms' costs in the long run (*i.e.*, $K'(z) \geq 0$). However, our interest is in costly regulation.

impact, *i.e.*, it will reduce the aggregate pollution, the spatial concentration of pollution, and the total pollution damage. In addition, it will reduce income for entrepreneurs in both regions (see (6)) because compliance with the standard results in an increase in firms' costs.

5.2. Uniform policy in the intermediate run

In the intermediate run, we must consider the impact of the regulation on firm relocation decisions. Because we are interested primarily in the range where partial agglomeration is the equilibrium outcome, we assume throughout that both before and after regulation z is in the range that yields this outcome. The total effect of the regulation in the intermediate run is a combination of the direct effect (given by the short run impacts) and the indirect effect through the impact on firm location decisions. The effect on location, in turn, has two components: an effect holding cost constant (see (18)) and an effect through its impact on fixed costs (see (17)). Moreover, we can show that for the environmental impacts the indirect effect of firm relocation dominates. As a result, instead of reducing aggregate pollution, with endogenous firm relocation the uniform regulation actually increases aggregate pollution, the spatial concentration of pollution, and total pollution damages.

Proposition 4. *In the intermediate run (with endogenous firm location but a fixed number of firms), under partial agglomeration* $\frac{d(Z_E^*+Z_W^*)}{dz}<0$, $\frac{d(Z_E^*-Z_W^*)}{dz}<0$, *and* $\frac{dTD^*}{dz}<0$, *i.e., an emission standard that restricts emissions per unit of output and raises fixed costs will increase aggregate pollution, the spatial concentration of pollution, and total pollution damages. Moreover, if the cost of the regulation is sufficiently high such that* $|K'(z)|>\frac{2(Y_E-Y_W)}{1-\varphi N(0.5+2\widetilde{\theta})}\left|\frac{\partial\lambda^*}{\partial z}\right|$, *the regulation also reduces aggregate income (i.e.,* $\frac{\partial Y_T}{\partial z}>0$).

Intuitively, if production causes less pollution as a result of the regulation, it becomes less costly for firms to agglomerate (i.e., for entrepreneurs to live in the region with more firms). As a result, the incentives for agglomeration will increase, and production will become more concentrated (i.e., λ^* will increase). This will increase aggregate demand (since entrepreneurial consumers will be moving to the location with lower prices), which will in turn lead to an increase in aggregate production and hence aggregate pollution. In addition, because the policy is assumed to be binding, compliance with the policy will result in an increase in firms' fixed costs. Ceteris paribus, this will further increase the benefit of agglomeration (since the impact on fixed costs will be lower in the larger region). Moreover, because a reduction in z increases both the total amount of pollution and its spatial concentration, it will result in larger pollution damages. Thus, rather than improving environmental quality and reducing the damages from pollution, if entrepreneurs move in response to the policy's impact on relative utility in the two regions, the regulation will be counter-productive in terms of overall environmental quality and pollution damages. (The environmental effects of an emission standard over the whole range of z are illustrated in Fig. 1.) In addition, if the cost of the regulation is sufficiently high, it will dominate the increasing agglomeration benefits, and the overall aggregate income will decrease.

It is important to note that the result that a uniform emissions standard can be counterproductive depends on several assumptions, including a) entrepreneurs care about local environmental quality, b) pollution is not perfectly mobile, and c) there exist agglomeration economies. If entrepreneurs do not care about local pollution, or if the pollution is a "global" public bad, a uniform emission standard will not affect firms' location decisions, and conventional results will prevail. In addition, when there are no transport costs and no co-location benefits, a uniform emission standard will have no effect on the relative benefit of locating in the two regions, and again the effect of the standard will be as expected. However, when these three conditions exist, the regulation will trigger an indirect effect through spatial relocation. Three forces cause the indirect effect of the regulation to dominate its direct effect (reflected in the short-run effect): 1) a uniform standard will lead to a larger environmental improvement in the more concentrated region (i.e., a larger reduction in pollution damages) due to the increasing marginal pollution damage ($\delta>0$), 2) as more firms move to the larger region, the co-location benefits in the region will increase, which will attract even more firms to the region, and 3) as more firms move to the large region, the overall prices of goods in the economy will decrease, which will lead to an increase in total demand and hence an increase in total production and pollution. Finally, it should be noted that, although the counter-intuitive results are derived under the specific model assumptions, the main driver of the results - adoption of cleaner technology makes agglomeration less costly - still holds even under more general model assumptions, implying that the increased spatial concentrations of firms that the regulation triggers through a reduction in agglomeration costs will offset at least part of the direct effect of the regulation, making it less effective.

5.3. Uniform policy in the long run

It is widely recognized in the literature that over time some firms may enter or exit the market in response to environmental regulation (Becker and Henderson 2000; List et al. 2003, 2004). We now consider how an emission standard affects firm entry/exit and the aggregate outcomes in the long run, again focusing on the region where partial agglomeration is a stable steady state. As shown in the proof of Proposition 3, in the long run,

$$\frac{\partial\lambda^*}{\partial z}<0 \tag{23}$$

$$\frac{\partial N^*}{\partial z}<0 \text{ if and only if } |K'(z)|<-\frac{(\varphi K+\Delta\pi)N^*}{2(1-\varphi\lambda^*N)\sqrt{\widetilde{\theta}}}\left.\frac{\partial\widetilde{\theta}}{\partial z}\right|_N. \tag{24}$$

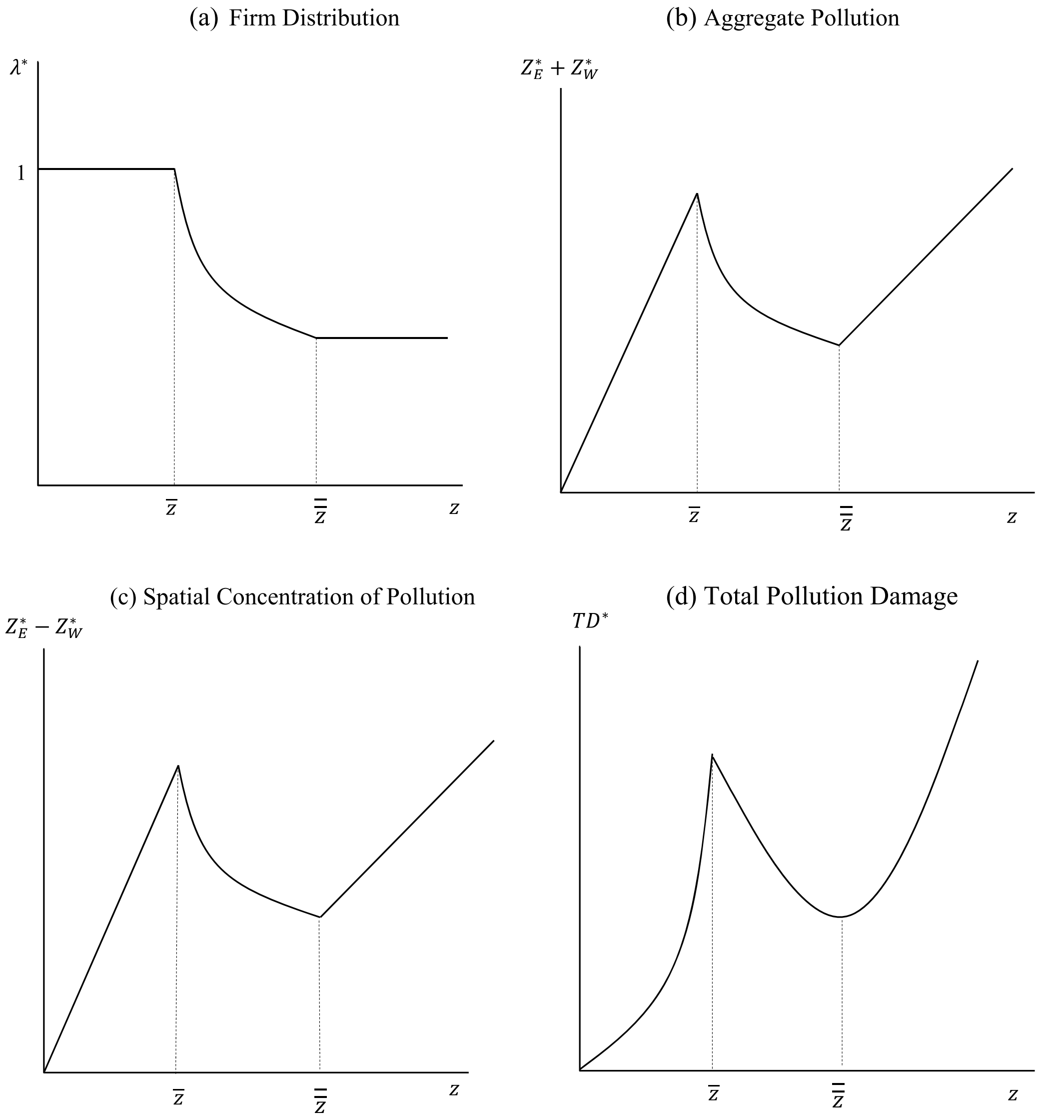

Fig. 1. The effect of an emission standard in the intermediate run when $\gamma > 0.5$.

Equation (23) suggests that in the long run, an emission standard will cause firms to be more concentrated in one region because the policy reduces the cost of agglomeration (by reducing pollution). This is similar to the effect seen in (18). However, the impact on the number of firms (*i.e.*, entry/exit) can be positive or negative, depending on the impact of the regulation on fixed costs. Intuitively, this effect has two components. On the one hand, the regulation will increase the spatial concentration of firms, which will increase the profit of entrepreneurs in the larger region and provides an incentive for more laborers to become entrepreneurs. On the other hand, the impact of the regulation on costs tends to discourage firm entry. The net effect on the number of firms depends on the relative magnitude of these two effects.

The effect of the standard on entry/exit will in turn determine how it affects environmental outcomes, but in the opposite way of what might be expected, namely, a regulation that induces entry will lead to lower pollution while a regulation that causes firms to exit will lead to more pollution.

Proposition 5. *In the parameter range where partial agglomeration is a stable steady state, in the long run, when $|K'(z)|$ is sufficiently small, imposing a uniform emissions standard will*

i) *increase aggregate income,*
ii) *increase the total number of firms in the industry,*
iii) *lower aggregate pollution, and*
iv) *if z is sufficiently high, reduce the spatial concentration of pollution and aggregate pollution damages.*

However, when $|K^{'}(z)|$ is sufficiently large, imposing a uniform emissions standard will

i) *reduce aggregate income,*
ii) *reduce the total number of firms in the industry,*
iii) *increase aggregate pollution, and*
iv) *if the emission intensity (z) is sufficiently high, increase the spatial concentration of pollution and aggregate pollution damages.*

Intuitively, when the impact of the regulation on firms' costs is sufficiently low, there is a "win-win" outcome. The increased benefit of agglomeration that results from the regulation will outweigh the low compliance cost, leading to higher income for entrepreneurs. The higher income for entrepreneurs will induce more laborers to become entrepreneurs, and hence more firms will be established. Firm entry will cause a smaller increase in pollution damage in the smaller region (because of its higher environmental quality and the strictly convex damage function). As a result, the smaller region will attract a larger share of the new entrepreneurs and both the firms and the population will be more spatially dispersed. This will, in turn, lead to lower total consumption (because of higher prices in the smaller region) and lower aggregate pollution. In addition, if the emission intensity is sufficiently high, the standard will also lead to lower spatial concentration of pollution and lower pollution damage. The opposite happens when the regulation significantly increases firms' costs (i.e., when $|K^{'}(z)|$ is sufficiently large).

The result that an environmental regulation will generate win-win outcomes in the long run when it has a small impact on firms' cost but it will have the opposite effect when it significantly increases firms' cost provides a rationale for government subsidies for emission reductions. Similarly, it could offer an argument for emission standards over emission taxes if the two policy instruments lead firms to adopt the same pollution control technology because the additional tax burden may cause some firms to exit in the long run and cause the surviving firms to be more spatially concentrated.

6. Effectiveness of differentiated environmental standards

So far, we have focused on the impact of a uniform regulation applied in both regions and found that, although such a policy is effective in the short run, it can be counter-productive in the intermediate and long run. This raises the question of whether a spatially differentiated policy that is applied only in the more polluted region (here the East) would be more effective.

6.1. *Differentiated policy in the short run*

To examine the impact of a differentiated policy, we expand the model in the previous section to allow emission intensities (and the associated fixed costs) to vary across the two regions. We consider a scenario where the region with more firms and hence more pollution (the East) has already put in place some regulation but still faces high pollution levels and is therefore considering additional regulation (a further reduction in z_E). Without loss of generality, we assume $z_E \le z_W$ (even though $Z_E > Z_W$) and hence $K_E = K(z_E) \ge K_W = K(z_W)$ and $K(z) \to +\infty$ as $z \to 0$. In addition, to keep the model tractable, we focus on the special case where $\tau = 0$ and $\delta = 0$.[21] Under these assumptions, the aggregate pollution and the difference in pollution between the two regions are given by:

$$Z_E + Z_W = \frac{(\alpha - w)MN}{2\beta}[(z_E - z_W)(\lambda - 0.5) + 0.5(z_E + z_W)] \tag{25}$$

$$Z_E - Z_W = \frac{(2\gamma - 1)(\alpha - w)MN}{2\beta}[(z_E + z_W)(\lambda - 0.5) + 0.5(z_E - z_W)]. \tag{26}$$

It is clear from (25) and (26) that, holding the total number of firms (N) and firm locations (λ) constant, a reduction in z_E will lead to the expected results in the short run, i.e., lower aggregate pollution, lower spatial concentration of pollution, and hence lower pollution damages.

6.2. *Differentiated policy in the intermediate run*

Our main interest is in the impact of the policy with endogenous firm location (and ultimately entry/exit). Again, we assume that entrepreneurs choose to locate in the region that gives them the highest utility. The utility difference for entrepreneurs located in the two regions is now

$$\Delta V \equiv V_E - V_W = -B(\lambda - 0.5 - \vartheta), \tag{27}$$

where

$$B = \frac{(\gamma - 0.5)MN(\alpha - w)}{\beta}(z_E + z_W) - \varphi N(K_E + K_W)$$

[21] In this case, agglomeration economies stem solely from co-location benefits, *i.e.,* $\varphi > 0$. Note that, unlike in the previous section where throughout $z_E = z_W$, here partial agglomeration can arise as a unique equilibrium even when $\delta = \tau = 0$.

$$\vartheta = \frac{1}{B}\left[\frac{(\gamma-0.5)MN(\alpha-w)}{2\beta}\ (z_W - z_E) - (1-0.5\varphi N)(K_E - K_W)\right].$$

As before, we are especially interested in the partial agglomeration equilibrium and assume if agglomeration occurs, it occurs in the East (*i.e.,* $\lambda^* > 0.5$). From (27), such a partial agglomeration can be a stable equilibrium if and only if $B > 0$, which we assume from now on. Under this assumption, we derive the following result, which defines the partial agglomeration equilibrium analogous to the one defined in Proposition 2.

Proposition 6. *Suppose $\gamma > 0.5$. There exists $(\overline{z}_E, z_E)$ such that, when $\overline{z}_E < z < z_E$, $\lambda^* = 0.5 + \vartheta$ is a unique stable equilibrium in the intermediate run with partial agglomeration, i.e., $\lambda^* < 1$.*

Now consider the effect of an increase in the stringency of the regulation in the East, i.e., a reduction in z_E. Proposition 6 implies that, within the range of partial agglomeration,

$$\left.\frac{\partial \lambda^*}{\partial z_E}\right|_N < 0 \quad \text{if and only if} \quad |K'(z_E)| < \frac{\lambda^*(\gamma-0.5)(\alpha-w)NM}{\beta(1-\varphi\lambda^* N)}. \tag{28}$$

This suggests that when the marginal cost of emission reductions is sufficiently small (i.e., $K'(z_E)$ is less negative), tightening the emission standard in the East will cause more, rather than fewer, firms to locate in the region. Intuitively, the effect of the policy depends on the relative magnitude of two effects. First, the unilateral policy reduces the pollution intensity and therefore the cost of agglomeration in the East. This tends to make the East more attractive to entrepreneurs. However, the additional costs associated with complying with the standard tends to make the East less attractive. When the first effect dominates (i.e., the condition on the right-hand side of (28) holds), the policy will cause more firms to locate in the East, and the short run effectiveness of the policy will be offset by the endogenous firm relocation.

Specifically, we can show the following:

Proposition 7. *Under partial agglomeration, in the intermediate run with endogenous firm relocation, an emission standard that restricts emissions per unit of output only in the East will reduce aggregate pollution. However, it will increase the spatial concentration of firms and pollution and total pollution damage if*

$$|K'(z_E)| < \frac{\lambda^*}{(1-\varphi\lambda^* N)}\left[\frac{(\gamma-0.5)(\alpha-w)NM}{\beta} - \frac{B(M+2N\vartheta)}{[2N\vartheta(z_E+z_W) - M(z_W-z_E)]}\right]. \tag{29}$$

Note that condition (29) is stronger than (28), which implies that when the cost of emission reduction is sufficiently low or the co-location benefits φ is sufficiently large, the increased agglomeration benefit from the regulation will dominate the compliance cost. This will lead to increased concentration of firms and pollution and larger pollution damage, even though total pollution is lower.

6.3. Differentiated policy in the long run

As in the case of a uniform policy, more stringent regulation in the East may also cause firm entry or exit, which may in turn affect the long-run steady state defined by the evolution of $\lambda(t)$ and $N(t)$. As before, we can show that the long run equilibrium depends on the pollution intensity.

Proposition 8. *Suppose $\gamma > 0.5$. There exists $(\overline{z}'_E, z'_E)$ such that, when $z_E \in (\overline{z}'_E, z'_E)$, the solution to the following two equations*:

$$\begin{cases} \lambda = 0.5 + \vartheta \\ N = H(0.5\overline{\pi} - (1-\varphi N\lambda)K(z_E)) \end{cases} \tag{30}$$

defines a unique long run stable steady state equilibrium (λ^*, N^*) with partial agglomeration, i.e., $\lambda^* < 1$. This is analogous to Proposition 3. Furthermore, analogous to (24), we have

$$\frac{\partial N^*}{\partial z_E} < 0 \text{ if and only if } |K'(z_E)| < \frac{\lambda^*(\gamma-0.5)(\alpha-w)NM}{\beta(1-\varphi\lambda^* N)}\ \frac{\varphi N K(z_E)}{(B+\varphi N K(z_E))}\ . \tag{31}$$

This suggests that a differentiated emission standard will induce firm entry if its impact on firms' fixed costs is sufficiently small or the co-location benefits φ is sufficiently large.

The overall long run impact of the differentiated policy that combines the effect in the intermediate run with the impact through entry/exit is summarized in the following:

Proposition 9. *In the parameter range where partial agglomeration is a stable steady state equilibrium, in the long run, if $|K'(z_E)|$ is sufficiently small, tightening the emission standard in the East unilaterally will*

i) *increase aggregate income,*
ii) *increase the total number of firms in the economy,*

iii) *lower aggregate pollution when the difference in emission intensity between the two regions is sufficiently small, and*
iv) *increase the spatial concentration of firms and total pollution damages.*

However, if $|K'(z_E)|$ *is sufficiently large, tightening the emission standard in the East unilaterally will*

i) *reduce aggregate income,*
ii) *reduce the total number of firms in the economy,*
iii) *increase aggregate pollution when the difference in emission intensity between the two regions is sufficiently large, and*
iv) *reduce the spatial concentration of firms and total pollution damages.*

Proposition 9 shows that, as with uniform standards across both regions, differentiated standards can also lead to unintended environmental impacts in the long run, depending on the cost of emission reductions, the level of co-location benefits, and the difference in emission intensity between the two regions. Specifically, when the cost of emission reduction is sufficiently high and the difference in emission intensity between the two regions is sufficiently large, tightening regulation in the more developed and polluted region will not necessarily reduce aggregate pollution. However, as with a uniform standard, a differentiated standard can also generate win-win outcomes in the long run, i.e., more manufacturing firms, higher aggregate income, lower aggregate pollution, if its impact on firms' fixed costs is sufficiently small.

6.4. *Optimal emissions standards*

The results above suggest that, when firms relocate in response to regulation, the regulation may not have the intended environmental benefit. In particular, increasing the stringency of an emission standard can actually increase pollution and pollution damages because of the indirect effect on firm relocation (see Propositions 7 and 9). This raises the question of how an optimal standard in the presence of relocation would compare to what would be prescribed by a traditional (short run) analysis where location is fixed. In this section, we explore this question, using the intermediate run analysis of the impact of a differentiated standard to illustrate the basic considerations underlying it. These stem from a recognition that there are three distortionary forces at work here: market power, pollution externalities, and agglomeration externalities. However, because the distortions due to market power and pollution both operate through output, there are only two distorted margins: output and location. Previous studies of agglomeration and pollution have considered the design of optimal policies when multiple instruments are available to control these different margins (*e.g.*, Henderson 1977; Arnott et al., 2008; Kyriakopoulou and Xepapadeas 2017; Regnier and Legras 2018). However, our primary interest here is in the role of environmental regulation when used by itself (as opposed to when coupled with other instruments). Thus, rather than discussing a first-best combination of policies, we focus here on a second-best optimal emissions standard.

Given the emissions intensity in the West, a second-best emissions standard in the East maximizes aggregate utility across both regions, given by

$$AV \equiv 0.5(M-N)V_E^L + \lambda N V_E^F + 0.5(M-N)V_W^L + (1-\lambda)NV_W^F \quad , \tag{32}$$

where V_E^L, V_E^F, V_W^L and V_W^F denote the indirect utility functions for laborers and entrepreneurs located in the East and West, respectively. Differentiating (32) with respect to z_E gives the following first-order condition for the optimal standard (see the derivation in the appendix):

$$\frac{\partial AV}{\partial z_E} = MD_{z_E} + \lambda N K_E'(z_E) + \frac{\partial AV}{\partial \lambda}\frac{\partial \lambda}{\partial z_E} = 0 \tag{33}$$

where $MD_{z_E} = 0.5MQ_E + 2(\lambda - 0.5)(\gamma - 0.5)NQ_E$ is the marginal pollution damage caused by increasing the emission intensity of firms in the East, $[-\lambda N K_E'(z_E)]$ is the marginal cost of reducing the emission intensity of firms in the East. Equation (33) suggests that when the spatial distribution of firms is optimal, i.e., when $\frac{\partial AV}{\partial \lambda} = 0$, we get the standard result that the optimal emission standard should be set such that the marginal benefit of reducing the emission intensity (given by the reduction in marginal damages) equals its marginal cost.

However, absent other location-based policies, the spatial distribution of firms will not generally be optimal. In this case, the optimal emission standard can be above or below the one set based on the conventional wisdom. Specifically, as shown in Proposition 6, when the marginal cost of reducing emissions intensity is sufficiently low, an emission standard in the East will increase the spatial concentration of firms in the region $\left(\frac{\partial \lambda}{\partial z_E} < 0\right)$, and the spatial distribution of firms in equilibrium under a standard set based on conventional wisdom will be more concentrated than the optimal level ($\frac{\partial AV}{\partial \lambda} < 0$ when $\Delta V = 0$) (see the proof in the appendix). In such cases, the optimal emission standard will be more stringent than the one set based on the conventional wisdom. The opposite is true when the marginal cost of reducing emissions intensity is sufficiently high and the spatial distribution of the firm is less concentrated than the optimal level. More generally, in the second-best setting, optimal standard will reflect a balancing of the impacts on the output/pollution and location distortions.

7. Empirical evidence: water pollution control in China

Our theoretical model implies that agglomeration economies will impact firm location decisions and ultimately the effectiveness of environmental regulation. In this section, we present initial empirical evidence in the context of water pollution control in China that both demonstrates the importance of agglomeration economies in determining the impacts of environmental regulation and supports our theoretical model specification and predictions. We first provide evidence that agglomeration economies affect the impact of regulation on both the number of firms and total emissions in a region, and that failure to account for this interaction can lead to inaccurate conclusions about the effects of regulation. We then provide evidence of a possible explanation for the link between agglomeration and regulatory impacts, namely, that firms in locations with higher agglomeration economies face lower costs of reducing emissions – a key assumption in our theoretical model. Our empirical application is based on an examination of the impact of regulation of chemical oxygen demand (COD) emissions from the pulp and paper industry in China.

7.1. Water quality regulation in China

With rapid economic growth, many rivers, lakes, and coastal waters in China have been severely polluted. A major strategy that the Chinese government uses to control water pollution is to designate severely polluted watersheds as “key regions” and then develop a “prevention-and-control plan” for each designated region (Wang et al., 2018). These plans define the government's goals on water quality and total emissions of the regulated pollutants for each key basin and set reduction targets for major pollution sources (*e.g.*, agriculture, industry, and domestic sources). To achieve the goals and targets, the government relies on a range of regulatory measures, including command-and-control approaches (*e.g.*, emission standards), economic incentives (*e.g.*, emission charges), and administrative measures (*e.g.*, shutting down small or inefficient paper mills). Thus, polluting firms in designated industries in these regions are subject to more stringent environmental regulation and higher public pressure for pollution control than firms in the non-designated regions.

Since 1996, the “three rivers and three lakes” (3R3L) basins have been incrementally designated as key regions for water pollution control. The basins of the three rivers (Huai, Liao, Hai) are located in northern China, while the basins of the three lakes (Tai, Chao, Dianchi) extend over the southern part of the country. The 3R3L basins cover 810,000 km^2, traverse 14 provinces, and are inhabited by around 360 million people. Table 1 documents the State Council's approval dates for the prevention-and-control plans for individual basins. In 1996, the Huai River basin was the first to be designated as a key region for water pollution control. The other five basins were designated in 1998 or 1999. The plan for each basin identifies the industries regulated to reduce major water pollutants. These industries are mainly at the 2-digit Chinese Industry Classification System level with some at the 3-digit or 4-digit levels.

Here we focus on COD emissions, which are the main water pollutant discharged by the manufacturing industries,[22] and on the largest COD-polluting industry, namely, pulp and paper (Chinese industry code (CIC) at 3-digit level: 221 and 222), which contributed 33% of the total COD emissions from the industrial sectors in China in 2002.[23] As shown in Table 1, not every basin regulates the pulp and paper industry, and those that regulate did not all start the regulation at the same time. Regulation of the pulp and paper industry provides us an ideal case study for several reasons. First, there is spatial and temporal variation in regulatory status of the counties. This variation allows us to identify the effects of regulation and its interaction with agglomeration economies. In addition, the counties differ significantly in the sizes of the pulp and paper industry and other industries. Finally, water pollutants such as COD can travel reasonably far downstream (Kahn et al., 2015), which is consistent with the case discussed in the theoretical section where there is some partial diffusion of pollution.

7.2. Data

Our empirical investigations rely on data for three years (1995 or 1998, 2004 and 2008) from three sources. The first is the Industrial/Economic Census. In particular, we use data from the China Industrial Census 1995, China Economic Census 2004, and China Economic Census 2008.[24] Information is included about each firm's identification code (*i.e.*, legal code), county of location, industry, birth year, ownership, employment, fixed assets, gross output, wage payments, and other related variables. Firms are classified following the 4-digit Chinese Industry Classification System.[25] We use this information to construct a measure of the number of firms in each county in each of the three years. In addition, based on the firm-level information, we construct measures of agglomeration economies for each county in our sample (see details below).

An advantage of the Census data is that the data include all manufacturing firms in China. A complete count of all firms is important when studying the impacts of environmental regulation on the total number of firms. The Census includes data on 2580 counties out of a total of 2609 (boundary-consistent) counties in the country.[26] The other 29 counties are not included in the Census because no

22 Regulatory efforts in the basins also aim to reduce phosphorus (*P*) and nitrogen (*N*) discharges into the lakes. The present paper focuses on the COD-emitting industries as the manufacturing sector is not a major source for TP or TN.

23 The pulp paper industry (Chinese industry code 22) contains three 3-digit industries: pulp (221), paper (222) and paper products (223). However, according to the government document, *Handbook on Emission Coefficients of Industrial Sources of Pollution for the First National Census on Pollution Sources*, the manufacturing of paper products is regarded as a less or even non-emitting sector. We therefore exclude those firms and include only firms in 221 and 222.

24 There were no censuses on industrial firms in other years during the 1995–2008 period.

25 In 2003, the Chinese Classification System was revised to incorporate more details for some sectors, while some other sectors were merged. Following Brandt et al. (2012), we construct a consistent classification so that the industry codes are comparable over the period.

26 We construct consistent area codes for counties in the data by using historical information on changes of county name and boundaries.

Table 1
Prevention-and-control plans for water pollution control in the 3R3L basins in China.

Key region	Approval date	Is the pulp and paper industry (CIC 221, 222) regulated?
Initial plan for the 9th Five-Year Period (1996–2000)		
Huai River	June 29, 1996	yes
Hai River	March 11, 1999	yes
Liao River	March 11, 1999	no
Tai Lake	January 6, 1998	yes
Chao Lake	June 7, 1998	no
Dianchi Lake	September 6, 1998	no
Plan for the 10th Five-Year Period (2001–2005)		
Huai River	January 11, 2003	yes
Hai River	March 4, 2003	yes
Liao River	January 30, 2003	yes
Tai Lake	August 31, 2001	no
Chao Lake	December 23, 2002	no
Dianchi Lake	March 12, 2003	no

Notes: 1. Data for this table are obtained from each basin's prevention-and-control plans, for the given periods, and the State Council's approval of these plans. 2. The 3R3L plans for the 11th FY (2006–2010) were approved by the central government in April 2008, which is beyond the periods covered by the present study. 3. The initial 1996–2000 plans for the three basins (i.e., Liao River, Chao Lake, and Dianchi Lake) emphasized reduction in emissions from agricultural and domestic sources and did not list any manufacturing industries for controlling COD emissions.

manufacturing activities are reported. Out of the 2580 counties, 623 are located in the 3R3L basins. As shown in Table 1, none of the prevention-and-control plans had been developed by 1995. In other words, the pulp and paper industry was not regulated in 1995 and we can therefore treat this as the "before regulation" period. The regulations were all put in place between 1996 and 2003. In 2004 and beyond, they covered all counties in the three river basins, which is 572 counties. Thus, these counties are treated as the regulated counties in the "post-regulation" period (2004 and 2008). Counties outside the 3R3L basins, as well as counties in the Tai, Chao, and Dianchi Lake basins are not regulated during these years.

The second data source is the Environmental Survey and Reporting (ESR) data administrated by the Ministry of Environmental Protection (or the former State Environmental Protection Administration). Beyond basic information such as name, industrial sector, location, and output value of the emitters, the ESR also includes emissions of major pollutants (*e.g.*, COD, ammonia nitrogen, sulfur dioxide) at the plant or firm level. Top firms/plants contributing 85% of total emissions of the major pollutants in a county are included in the ESR. Wang et al. (2018) provide details about the ESR system and its way of collecting and verifying data. We use these data to construct county-level COD emissions from the pulp and paper industry for each county in each of the years in our sample.

The third data source is the Annual Surveys of Industrial Firms (ASIF) conducted by the National Bureau of Statistics of China from 1998 to 2007. The ASIF data have been employed widely by studies on firm behavior, productivity and economic growth (see, for example, Brandt et al., 2012). The data contain similar firm-level information as the economic census. While its frequency is higher, one key difference between the ASIF and the Census datasets is that the ASIF does not cover non-state-owned (non-SOE) firms with sales of less than 5 million RMB per year (about US $800,000). We use these data in our investigation of the impact of agglomeration on compliance costs. We clean the ASIF data by following Brandt et al. (2012). In particular, we use their procedures to construct price indexes and the variables for physical capital stock.

7.3. Impact of regulation on number of firms and emissions

In this section we present initial empirical evidence regarding the role of agglomeration economies in determining the effect of regulation on the total number of firms and the total COD emissions from the pulp and paper industry.

7.3.1. Methodology

The empirical methodology is similar for both analyses. The only difference between the two analyses is the outcome variable. For example, we estimate the following model to evaluate the effect of regulation, agglomeration economies and their interaction on the number of firms[27]

$$log(N_{ct}) = \alpha_1(Post_t \times Regulate_c) + \alpha_2 \times AE_{ct} + \alpha_3[(Post_t \times Regulate_c) \times AE_{ct}]$$

[27] Our theoretical model suggests that environmental regulation may have spillover effects, i.e., it may also affect the total number of firms in non-regulated counties. The presence of spillover effects violates the Stable Unit Treatment Value Assumption (SUTVA) for the Difference-in-Difference (DD) model used here. Although this does not affect any large sample property of the DD estimator such as consistency and asymptotic normality, in the presence of spillover we must interpret the coefficient on regulation as the difference between the impact of the regulation in regulated counties and the impact in non-regulated counties (see Zhou et al., 2017). The year-fixed effect in the model picks up the spillover effects.

$$+\upsilon_c + \delta_t + \beta_1(Trend_t \times Regulate_c) + u_{ct}. \tag{34}$$

In this model, N_{ct} is the number of firms in county c and year t. We construct it from the Census data. $Post_t$ is an indicator variable equal to 0 for year 1995 and 1 for years 2004 and 2008. $Regulate_c$ is also an indicator variable equal to 1 if county c regulates the industry in the post-regulation years and zero otherwise. υ_c *and* δ_t are the county- and year-fixed effects, and u_{ct} is an error term. We include county-level fixed effect υ_c to control for unobservable county attributes that are time-invariant. The time fixed effect indexed by δ_t absorbs year-specific unobservables. We also add a linear regional trend, $Trend_t \times Regulate_c$, to control for the region-specific time-varying unobservables that affect county outcomes.

A key explanatory variable in (34) is AE_{ct}, which is our measure of agglomeration economies in county c in year t, proxied by county-level manufacturing employment outside of the pulp and paper industry (see further discussion below). The urban and regional literature finds that both localization economies (own-industry employment) and urbanization economies (other-industry employment) are important measures of agglomeration economies. While many studies on the determinants of firm births and employment consider both measures and find that localization economies are more important than urbanization economies at the margin (*e.g.*, Henderson, 2003; Rosenthal and Strange, 2003), there is much evidence supporting the opposite conclusion (*e.g.*, Glaeser et al., 1992). Unfortunately, the limitation of the Census data prohibits us from including a variable representing localization economies. For the analyses on the number of firms in the counties, it is inappropriate to use same-year own-industry employment (localization economies) in the counties to explain it. A normal practice in the literature is to measure localization economies in an earlier year and use it in a regression model like this. However, the Censuses cover only three years with different gaps between them (1995, 2004, and 2008). A potential data source for constructing the localization economies variable is the ASIF dataset. However, as noted above the ASIF covers only SOEs and large non-SOEs. If we use these data to measure localization economies for the industry, severe measurement errors exist. For instance, employment in the ASIF manufacturers account for only about 78% of actual total manufacturing employment in 2004. The percentage also varies significantly across counties. The census includes 1617 counties having the presence of the pulp and paper industry in 2004, but only 981 of them can be found in the ASIF. The data sample would shrink significantly if we use the ASIF data for the county-level analysis.[28] Therefore, we instead use the census data and focus on the role of urbanization rather than localization economies.

The primary obstacles in estimating the interaction effect (α_3) are sample selection bias and simultaneity and omitted-variable biases. Sample selection bias occurs if the treatment group (i.e., regulated counties) and control group (i.e., non-regulated counties) significantly differ in their pretreatment characteristics. Simultaneity bias may arise because a county's agglomeration patterns are both cause and consequence of firms' location choices. Omitted-variable bias could also exist in ordinary least square (OLS) estimates. For example, agglomeration is a result of regional growth, and thus potentially correlated with socioeconomic confounders, such as income, production and consumption amenities, and human capital, which are also important determinants of firm births and deaths.

To mitigate the endogeneity biases, we rely on an identification strategy that has two elements. First, when estimating the impact of regulation, we use propensity score matching, which allows us to construct a comparison group with counties not regulated by the 3R3L policy to serve as the counterfactuals for the treatment group with regulated counties. This approach pairs each regulated county with an unregulated county based on certain observable attributes. We conduct a balancing test to ensure that the unregulated county can serve as the counterfactual for the regulated county. (See further details below.)

Second, in addition to baseline OLS regressions, we include regressions that instrument for the agglomeration economics variable to address the potential endogeneity issue with this variable, following the approach used by Bartik (Bartik, 1991; Goldsmith-Pinkham et al., 2020). To construct the instrumental variable, we forecast manufacturing employment by county and industry based on employment in an early year at the two-digit CIC level and the national growth rates in individual industries. For example, the employment of industry i in county c and year t is forecasted as

$$L_{ic,t-\tau} * \frac{\sum_c L_{ic,t}}{\sum_c L_{ic,t-\tau}},$$

where the first term denotes employment in an early year. The second term represents the national growth rate of this industry between the two years. Summing the forecasted numbers of all industries except the pulp and paper industry, we obtain the value of the instrumental variable.

Equation (34) differs from specifications in early empirical studies (*e.g.*, Becker and Henderson, 2000) in that it includes the interaction term between environmental regulation ($Post_t \times Regulate_c$) and agglomeration economies (AE_{ct}). For a county that regulates the industry ($Regulate_c = 1$), the effect of the regulation, *i.e.*, going from $Post_t \times Regulate_c = 0$ to $Post_t \times Regulate_c = 1$, equals

$$\Delta log(N_{ct}) = \alpha_1 + \alpha_3 \times AE_{ct}, \tag{35}$$

where α_1 measures the direct effect of the regulation, and $\alpha_3 \times AE_{ct}$ measures the interaction effect.

[28] Nonetheless, as a check, we construct the localization economies measure and re-estimate an extended model that includes the effect of localization economies and its interaction with the regulation variable. The results show that the interaction effect between regulation and localization economies on firm numbers is statistically insignificant.

7.3.2. Results: impacts on the total number of firms

As noted above, we use propensity score matching to construct a comparison group with counties not regulated by the policy to serve as the counterfactuals for the treatment group with regulated counties. The county-level covariates for matching include the logarithms of the number of firms, manufacturing employees, own-industry employees and wage. We construct the comparison group by one-to-one matching. (For a description of the matched data sample, see Panel A of Table A1 in the appendix.) Table 2 presents the balancing test results for the control and treated groups. It suggests no significant differences between the two groups in all covariates used and even those not used in the matching approach. These results suggest that our matching strategy performs well in extracting reasonable comparison counties, similar to the regulated counties.

Using the matched sample, we estimate model (34) to quantify the effects of regulation and its interaction with agglomeration economies. Column 1 of Table 3 presents OLS results from a widely used specification that ignores the interaction effect between regulation and agglomeration economies (*e.g.*, Becker and Henderson. 2000; List et al., 2004). It suggests that the water quality regulation has a statistically insignificant effect on the number of firms in the regulated industries, while agglomeration economies have a positive and statistically significant effect. Columns 2 and 3 report estimates that include the interaction term between regulation and agglomeration economies. The model in Column 3 controls for the possibility of differing preexisting trends in the regulated counties and non-regulated counties, while the model in Column 2 does not. Nonetheless, both specifications give essentially the same results. After the interaction term is included, both the direct effect of the regulation and the interaction effect with agglomeration economies become highly statistically significant. The direct effect of the regulation is negative and significant, while the interaction effect is positive and significant. This is consistent with the predictions of our theoretical model, namely, that agglomeration economies have an indirect effect on the impact of environmental regulation that works in the opposite direction of the direct effect. In other words, the regulation will have less of a negative impact on economic activity in a county with greater agglomeration economies.

As noted above, agglomeration economies could be endogenous. Column 4 in Table 3 reports the results when we address potential endogeneity by instrumenting for agglomeration economies using a Bartik approach as discussed earlier. This is our preferred specification. The estimation results again show that agglomeration economies work to offset, at least partially, the negative effect of regulation on the number of firms. Moreover, the results imply that the regulation would actually increase the total number of firms in the pulp and paper industry in a county with sufficiently large agglomeration economies. In fact, equation (35) and the coefficient estimates in Column 4 of Table 3 suggest that, if a regulated county has more than 27,600 manufacturing workers outside the industry, the regulation would lead to an increase in the number of firms in the industry.[29] There are 199 counties (out of 572 regulated ones) in our sample that meet such a threshold for agglomeration economies in 2004.

To check the robustness of our estimates, we conduct a number of analyses and report the estimated coefficients and confidence intervals in Fig. 2. (For comparison purposes, Row 1 plots our main results from Column 4 of Table 3.) First, the literature lacks consensus about the inclusion of variables for the matching step. While it is safe to include a large number of covariates in the step, doing so is likely to lead to fewer matched pairs. Therefore, we choose a fair number of variables to ensure close similarity between the treated and control counties while keeping the highest number of matched pairs. Table 2 suggests that our baseline matching strategy performs very well. Nevertheless, we try different sets of county-level covariates for matching. As an example, the second row of Fig. 2 summarizes the regression results when we use the logarithms of the number of non-SOE firms, manufacturing employees, manufacturing output, and own-industry output as the variables.

The second check addresses the concern that some contemporaneous water quality policies may confound our estimates. We conduct a thorough review of possibly related national or regional environmental policies during the period and believe that the most likely candidate is the efforts to limit emissions from the industry in two major paper-producing provinces, namely, Shandong and Henan. In 2003, these two provinces tightened their own regulations on this industry by implementing stricter provincial emission standards. To address this concern, we add a province by year FE for either province into the model. The third row of Fig. 2 summarizes the results of this version of the model.

Firms may benefit from agglomeration economies not just in their own counties, but also in neighboring counties. To examine this, instead of measuring the level of agglomeration economies based on the number of manufacturing employees outside the own industry in the county only, we measure it using manufacturing employment outside the industry in the prefecture (which consists of multiple counties). The results are presented in the fourth row of Fig. 2.

As the fourth robustness check, we drop all unregulated counties from the provinces that contain no regulated counties. In other words, we limit our sample to provinces that contain the 3R3L basins. Thus, the sample becomes smaller. Row 5 of Fig. 2 reports these results.

The unit of observation in the Census is the firm, not plant. It is possible that an emitting firm from a regulated county has plants in non-regulated counties. However, we have no information about the distribution of those plants. Ignoring that could bias the estimates. Fortunately, about 98.6% of all observations in the pulp and paper industry in 2004 are single-plant firms. The results when we limit the sample to one-plant firms are presented in Row 6 of Fig. 2.

In the last robustness check, we use $log(1 + number\ of\ firms)$ as the dependent variable, so some county-year observations with no polluting firms also enter the sample. The last row in Fig. 2 presents the results.

[29] More specifically, the numbers suggest that, if $AE_{ct} \geq 3.318$, then $\Delta log(N_{ict}) \geq 0$. Because AE_{ct} is defined as the logarithm of the number of other-industry employees (in thousand), the condition $AE_{ct} \geq 3.318$ implies that the county has more than 27,600 manufacturing workers outside the industry.

Table 2
Balancing test: Sample for the firm number regressions.

Variables	Matched sample			Unmatched sample		
	treated	control	p value	treated	control	p value
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Covariates Used in Matching						
Number of firms, 1995	1.094	1.091	0.947	0.998	0.900	0.023
Employees, all industries, 1995	9.901	9.965	0.492	9.742	9.395	0.000
Employees, 1995	5.193	5.170	0.856	5.194	4.881	0.003
Wage, 1995	1.239	1.248	0.814	0.272	0.159	0.008
Panel B: Covariates Not Used in Matching						
Number of new firms, 1995	0.278	0.221	0.369	9.507	9.172	0.001
Output value, 1995	9.659	9.641	0.879	13.878	13.479	0.000
Output value, all industries, 1995	14.021	13.909	0.232	0.924	0.822	0.020
Number of non-SOE firms, 1995	1.011	0.969	0.461	1.269	1.445	0.000
# counties	419	419	–	531	1549	–

Notes: All attributes used in the matching approach are historical records in 1995 during the pre-regulation period. All attributes are in natural logarithms. For the matched sample, Columns (1) and (2), respectively, report the sample means of covariates for the treatment and matched control groups. Column (3) reports p values from testing the null hypothesis that there is no difference between the two groups. The next three columns are for the unmatched sample. Total number of counties in the table (i.e.,2080) is smaller than total number of counties in the Census because we remove those counties that have no firms in the industry during the whole 1995–2008 period.

Table 3
The interaction effect of water quality regulation and agglomeration economies on the total number of firms.

	Dependent variable: *log*(*number of firms*)			
	OLS (1)	OLS (2)	OLS (3)	2SLS (4)
$Post_t \times Regulate_c$	−0.062 (0.070)	−1.240*** (0.158)	−1.189*** (0.159)	−1.513*** (0.237)
AE_{ct}	0.142*** (0.043)	0.110** (0.043)	0.108** (0.043)	0.320*** (0.085)
$(Post_t \times Regulate_c) \times AE_{ct}$		0.353*** (0.045)	0.375*** (0.046)	0.456*** (0.065)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Control for preexisting trends	N	N	Y	Y
N	2221	2221	2221	2221
KP F-statistic	–	–	–	152.50

Notes: The level of agglomeration economies (*AE*) is measured by the logged value of employment (in thousand) outside the own industry. In column 4, we instrument for agglomeration economies using a Bartik approach. See the text for the construction of the instrument variable. Standard errors in parentheses. The standard errors are clustered at the county level. ***, **, and * denote significant at 1%, 5% and 10%, respectively.

In all cases, these robustness checks confirm the baseline conclusion that, as predicted by our theoretical model, the direct effect of the regulation reduces the number of firms in the county, but that this direct effect is offset by the interaction effect with agglomeration economies. Thus, regulation has dampens economic activity less (and in some cases may even enhance it) in counties with large agglomeration economies. This main conclusion is robust to potential threats from alternative set of covariates for matching, confounding policies, alternative definition of key variables, and data sample.

7.3.3. Results: impacts on the total emissions

We next evaluate the role of agglomeration economies in determining the effectiveness of the regulations (*i.e.*, their impact on total emissions) by estimating model (34) using *log*(*total emissions*) as the dependent variable. For this analysis, we aggregate firm-level COD emissions from the ESR data to obtain total emissions from the industry for each county in the years (1998, 2004 and 2008). Here, we use the year 1998 instead of 1995 as the pre-policy period because 1998 is the first year that the emission data can be obtained from the ESR. Due to this data limitation, we have to remove those counties that were designated as the regulated counties in or before 1998. As indicated by Table 1, those counties are located in the basins of Huai River and Tai Lake. The removal shrinks our data sample. In addition, the ESR did not cover as many counties as the Census. Thus, the sample size for investigating the impacts on total emissions is much smaller than that for the firm number regressions. In addition, agglomeration measures for 1998 are obtained from the ASIF data (For 2004 and 2008, they are obtained from the census data.).

As for the analysis of the number of firms, we first apply propensity score matching to obtain a suitable control group. The covariates for the matching step are total emissions, own-industry output value, total output value of the manufacturing sector, and wage (Panel B of Table A1 in the appendix provides a description of the matched data sample.). Table 4 presents a balancing test for the obtained control and treatment groups. As before, the results indicate that the matching performs well in generating a suitable control group.

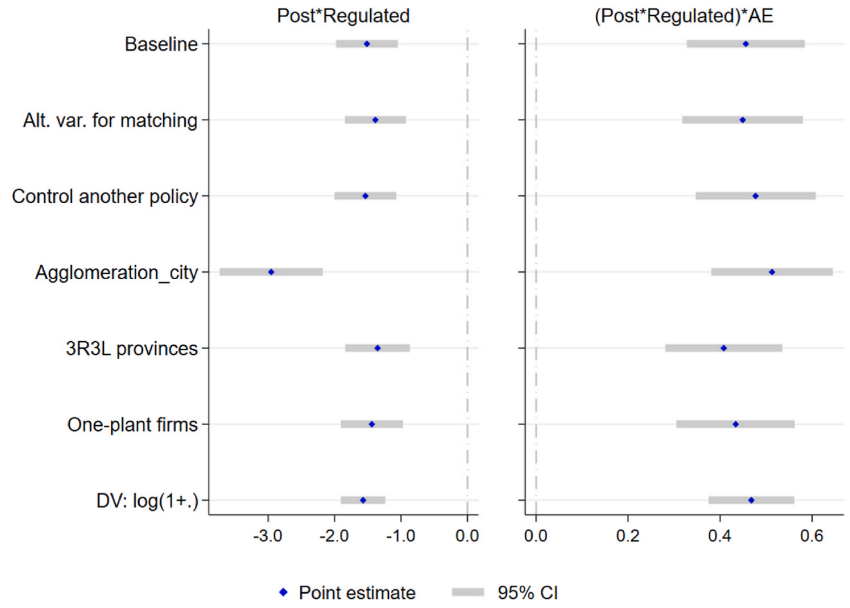

Fig. 2. Robustness checks: The interaction effect of water quality regulation and agglomeration economies on the total number of firms.

Table 4

Balancing test: Sample for the total emissions regressions.

Variables	Matched sample			Unmatched sample		
	treated	control	p value	treated	control	p value
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Covariates Used in Matching						
Total emissions, 1998	13.272	12.605	0.391	13.207	12.957	0.596
Output value, 1998	11.160	11.165	0.987	11.163	10.772	0.104
Output value, all industries, 1998	14.457	14.237	0.452	14.395	14.508	0.686
Wage, 1998	1.858	1.876	0.834	1.848	2.189	0.000
Panel B: Covariates Not Used in Matching						
Employees, 1995	6.976	6.800	0.460	9.762	9.824	0.794
Employees, all industries, 1995	9.813	9.704	0.674	6.982	6.365	0.005
Number of firms, 1998	1.208	1.079	0.511	1.244	0.795	0.002
# counties	34	34	–	36	188	–

Notes: All attributes used in the matching approach are historical records in 1998 during the pre-regulation period. All attributes are in natural logarithms. For the matched sample, Columns (1) and (2), respectively, report the sample means of covariates for the treatment and matched control groups. Column (3) reports p values from testing the null hypothesis that there is no difference between the two groups. The next three columns are for the unmatched sample.

Table 5 reports the results regarding the impact of regulation on total emissions. Similar to what we found for the impact of regulation on the number of firms, in the OLS specification that ignores the interaction effect of agglomeration and regulation, the effect of the regulation is insignificant. However, when the interaction effect is added (in Columns 2–4), the coefficients representing the direct effect of the regulation become statistically negative and the interaction effect is statistically positive. All of these results, including the 2SLS results that account for endogeneity of agglomeration economies, suggest that the water quality regulation has a direct and an interaction effect on the total emissions. Again, this is consistent with the predictions of our theoretical model, namely, that agglomeration economies have an indirect effect on the impact of environmental regulation that works in the opposite direction of the direct effect. As a result, regulation reduces pollution by less in counties that have more other-industry employment. When agglomeration economies are sufficiently large in a county, regulation might actually increase total emissions, rather than decrease them. According to the estimated coefficients reported in the last column of Table 5, if a regulated county has more than 26,100 manufacturing workers outside the industry, the regulation would lead to an increase, rather than decrease, in total emissions from the pulp and paper industry. About one third of the regulated counties in our sample (211 out of 572 regulated ones) met this threshold for agglomeration economies in 2004.

We run a number of regressions to check if our results are robust to potential threats from alternative set of covariates for matching (total emissions, own-industry employment, total manufacturing employment, and number of firms), confounding policies, alternative

Table 5
The interaction effect of water quality regulation and agglomeration economies on total emissions.

	Dependent variable: *log*(*COD emission*)			
	OLS (1)	OLS (2)	OLS (3)	2SLS (4)
$Post_t \times Regulate_c$	−0.399 (0.755)	−3.622** (1.517)	−3.202** (1.472)	−4.784*** (1.785)
AE_{ct}	−0.145 (0.697)	−0.158 (0.697)	−0.223 (0.702)	0.314 (1.315)
$(Post_t \times Regulate_c) \times AE_{ct}$		0.937** (0.425)	1.023** (0.434)	1.466*** (0.472)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Control for preexisting trends	N	N	Y	Y
N	185	185	185	182
KP *F*-statistic	–	–	–	20.30

Notes: The level of agglomeration economies (*AE*) is measured by the logged value of employment (in thousand) outside the own industry. In column 4, we instrument for agglomeration economies using a Bartik approach. See the text for the construction of the instrument variable. The Stock-Yogo critical value is 4.58 at the 15% level. Standard errors in parentheses. The standard errors are clustered at the county level. ***, **, and * denote significant at 1%, 5% and 10%, respectively.

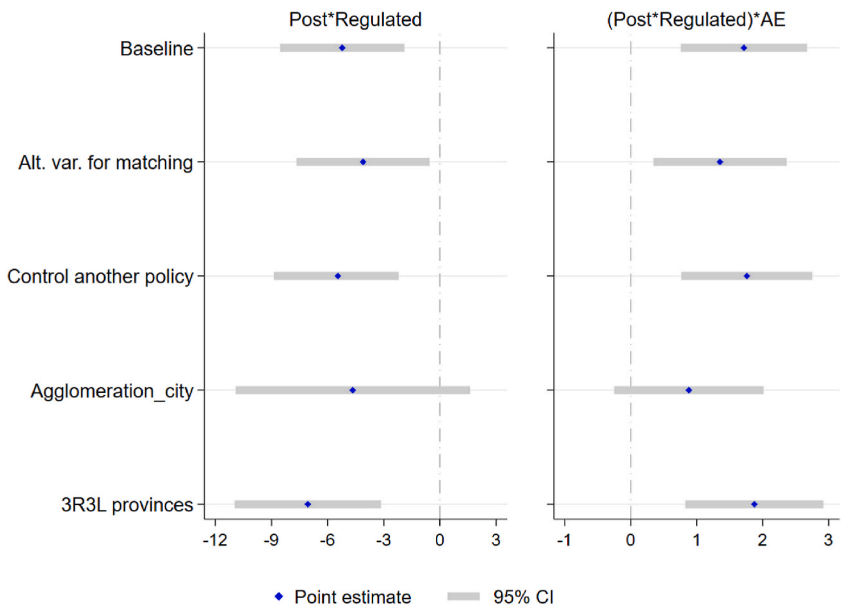

Fig. 3. Robustness checks: The interaction effect of water quality regulation and agglomeration economies on total emissions.

definitions of the agglomeration economies variable, and different data samples. The procedures for doing these checks are similar to the ones used in the above analysis on the number of firms. Fig. 3 summarizes the results from the robustness checks for the total emissions regressions. We also apply model (34) to the unmatched data sample and obtain qualitatively similar results. Tables A2 and A3 in the appendix report the results regarding the effect on total number of firms and total emissions, respectively.

The estimated coefficients from our preferred specification (reported in the last column of Table 5) can be used to generate an estimate of the overall magnitude of the impacts of COD regulation for the pulp and paper industry that incorporates both the direct and indirect effects through agglomeration. Our model suggests that, if a regulated county c in year t had not been regulated, the logarithm of its emissions in that year would have been:

$$log(\widetilde{EM}_{ct}) = log(Em_{ct}) - \alpha_1(Post_t \times Regulate_{ct}) - \alpha_3[(Post_t \times Regulate_{ct}) \times AE_{ct}], \tag{36}$$

where Em_{ct} is actual emissions level and $\widetilde{EM}_{ct} = Em_{ct}$ for unregulated counties. For each county, we can calculate the change in emissions induced by the regulation (including both the direct effect and the interaction with agglomeration economies) as $Em_{ct} - \widetilde{EM}_{ct}$. To get a sense of the effects in relative terms, we calculate the ratio between the sum of the induced changes and the sum of the emissions under the counterfactual of no regulation as:

$$\frac{\sum_c (Em_{ct} - \widetilde{EM}_{ct})}{\sum_c \widetilde{EM}_{ct}} \times 100\%.$$

Obviously, the ratio depends not only on the effects for individual counties, but also on the distribution of industrial employment among the counties. Our calculation shows that across our sample, while actual emissions from the industry were 935.45 metric tonnes in 2004, the number would have been 771.28 if no regulations had been implemented. This implies that, because of the effect of agglomeration economies, the regulations actually increased total emissions from the industry across all counties in our sample by about 21% in 2004.[30] However, it should be noted that many counties with lower COD emissions or those with a small pulp and paper industry are under-represented in the ESR dataset. Most of such counties are unregulated. According to the official *China Environment Yearbook*, total emissions from the 2-digit industry (CIC 22) in 2004 were 1488.26 metric tonnes. A simple comparison suggests that about 553 metric tonnes (=1488.26–935.45) of COD emissions are not recorded in the ESR data. Assuming that all of the 553 metric tonnes of emissions were generated by firms from unregulated counties, we would then estimate that as a result of agglomeration economies the regulations actually increased emissions by about 12%.[31] This implies that the potential for agglomeration economies to undermine, and possibly completely offset, the benefits of regulation is not only theoretically possible (as shown by our theoretical model) but can also arise in a real-world policy context.

7.4. Evidence regarding effect of agglomeration on pollution control costs

Our theoretical model assumes that agglomeration economies reduce the cost of compliance with regulation because regulated firms in an agglomerated region are more likely to have access to information, skilled labor, and technology needed for compliance. In this section we provide evidence supporting this assumption.

For this analysis, we link the ASIF and the ESR datasets by using each firm's legal code and name. Table 6 provides summary statistics of the sample obtained from merging these two datasets.[32] The growth rates for both input and output, approximated by $\Delta log(x) \approx \frac{\Delta x}{x}$, varied significantly among the firms in the industry. On average, from 1998 to 2007 the output value for the pulp and paper industry grew by 6.9% annually, while the total emissions decreased by 8.4%. In addition, the level of agglomeration economies varied significantly among counties that host firms in the industry.

To examine the impact of agglomeration on the costs of pollution control, we estimate a firm-level production function that considers COD emissions as one of the input factors required to produce the desirable output (Copeland and Taylor, 2013). Under this setup, the cost of COD emission reduction can be measured by the opportunity cost in terms of loss in the desired output. Following previous studies on estimating production functions (*e.g.*, Haskel et al., 2007), we estimate the following difference model for firms from a polluting industry:

$$\begin{aligned}\Delta logY_{ft} &= \rho_k \Delta logPK_{ft} + \rho_l \Delta logL_{ft} + \rho_m \Delta logI_{ft} + \rho_z \Delta logZ_{ft} + \rho_{za} \times \left(\Delta logZ_{ft} \times AE_{ct-1}\right) \\ &+ \rho_a AE_{ct-1} + \rho_x X_{ft-1} + \eta_j + u_t + v_f + \varepsilon_{ft},\end{aligned} \tag{37}$$

where subscripts f, j, c and t index the firm, 4-digit sub-industry within the polluting industry, county, and year, respectively; Δ denotes the change between years $t-1$ and t; Y is the output value; PK and I are, respectively, the values of physical capital and intermediate inputs; L is labor input; Z is the level of COD emissions; X_{ft-1} is a vector of firm characteristics in year t-1 (i.e., ownership types; firm age indicators on whether the firm is aged 10+ years; exporting status indicator on whether the firm exports; a dummy on whether the firm has multiple plants); and ε_{ft} is an error term. The model also controls for a set of fixed effects: industry shocks with 4-digit industry fixed effects (η_j), temporal fluctuations with year fixed effects (u_t), and firm fixed effects (v_f). The time-differencing equation controls for time-invariant unobservables that may affect output.

To examine the impact of agglomeration economies on the cost of COD emission reduction, we include AE_{ct-1}, which is the level of agglomeration economies in year $t-1$, both directly and interacted with the change in COD emissions. The elasticity of output value with respect to COD emissions is then given by $\rho_z + \rho_{za} AE_{ct-1}$. This elasticity measures the opportunity cost of COD emission reduction (in terms of loss in output value) under the existing technology.

Table 7 reports the estimation results from applying model (37) to the constructed data sample. In the first set of results, reported in Columns 1 and 2, we measure agglomeration economies using total employment outside the pulp and paper industry. In Column 1 we present OLS estimates. However, the measures of agglomeration economies could be endogenous even if they are based on the total employment or output value outside the own industry in the previous year because firms from other industries could be attracted to a county with a fast-growing industry to enjoy agglomeration benefits. To address this potential endogeneity issue, in Column 2 we present estimates that instrument for agglomeration economies again using a Bartik approach. Specifically, we use employment in an early year (1995) at the two-digit CIC level and the national growth rates in each industry to forecast manufacturing employment in the industry and county.

The results in Columns 1 and 2 of Table 7 suggest that COD emissions are one of the "input factors" required to produce the desirable output in the pulp and paper, *i.e.*, COD emission reductions in the industry can only be achieved at a cost of lower output under the existing production technology. For the pulp and paper industry, this opportunity cost of emission reduction is lower for

[30] This is calculated as $\frac{935.45-771.28}{771.28} \times 100\%$.
[31] This is calculated as $\frac{935.45-771.28}{771.28+553} \times 100\%$.
[32] A shortcoming of ESR is that it reports emissions at the plant level, and if a firm has multiple plants, some of the plants may not be required to report their emissions. Fortunately, Census information suggests that more than 95% of all observations in our sample are single-plant firms. In addition, if the output values of a firm from the two datasets differ by more than 10%, the firm was excluded from the final merged sample.

Table 6
Summary statistics: The merged economic data (ASIF) and environmental data (ESR).

Variable	Mean	Std. Dev.	Min	Max
$\Delta log(output\ value)$	0.070	0.219	−0.877	0.831
$\Delta log(labor)$	−0.020	0.159	−0.783	0.725
$\Delta log(capital\ stock)$	0.051	0.256	−1.163	1.670
$\Delta log(intermediate)$	0.069	0.251	−1.311	1.504
$\Delta log(COD\ emission)$	−0.088	0.808	−2.880	2.982
Agglomeration economies: employment	3.384	1.361	−1.784	7.731
Agglomeration economies: output value	8.723	1.700	3.143	13.202
Operating 10+ years	0.534	0.499	0	1
Having multiple plants	0.072	0.258	0	1
Export status	0.158	0.365	0	1

Notes: 1. The firms are from the pulp and paper industry (CIC 221, 222). N = 933; 2. The first five rows are for one-year change in the output, input, and emission variables from 1998 to 2007, respectively. 3. The other variables are measured at year t-1. Agglomeration economies based on employment is logged value of local manufacturing employment outside the own industry (in thousand). Agglomeration economies based on output value is the logged value of local manufacturing output value outside the own industry (in million (2007) RMB *yuan*).

Table 7
Agglomeration economies and marginal cost of abatement.

	Dependent variable: $\Delta log(Output\ value)$			
	AE based on employment		*AE* based on output value	
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
$\Delta log(COD\ emission)$	0.061**	0.044*	0.133**	0.108**
	(0.026)	(0.024)	(0.056)	(0.052)
$\Delta log(COD\ emission) \times AE_{ct-1}$	−0.018**	−0.012*	−0.015**	−0.012**
	(0.008)	(0.007)	(0.007)	(0.006)
Other controls	Yes	Yes	Yes	Yes
4-digit industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	933	929	933	929
KP *F*-statistic	–	24.942	–	25.207

Notes: 1. There are twelve unique firms that ever relocated across counties during the period. The regressions control for county fixed effects. 2. Standard errors in parentheses. The standard errors are clustered at the county level. ** and * denote significant at 5% and 10%, respectively.

firms located in counties with larger agglomeration economies. For example, in a county with 1000 local manufacturing workers outside the own industry (*i.e.,* $AE_{ct-1} = ln(1) = 0$), a 1% reduction in COD emissions will lead to a 0.06% reduction in output values in the pulp and paper industry. This opportunity cost of emission reduction is almost twice as high as that in a county with 5000 local manufacturing workers outside the own industry (*i.e.,* $AE_{ct-1} = ln(5) = 1.609$), where a 1% reduction in COD emissions will lead to only a 0.032% reduction in output values in the pulp and paper industry.

Because ASIF does not include non-state-owned firms with annual sales of less than 5 million RMB, using employment may lead to an underestimate of agglomeration economies. In 2004, employment in the ASIF manufactures account for only about 70% of total manufacturing employment in China. However, the ASIF firms produce more than 90% of the total industrial output in China. Therefore, a measure of agglomeration economies based on output value may better reflect actual levels. We report estimates using an output-based measure of agglomeration economies in Columns 3 and 4 of Table 7. These results are similar to those obtained from the employment-based measure of agglomeration economies and confirm that the opportunity cost of COD emission reduction is lower for firms located in counties with larger agglomeration economies. In another exercise, we run similar regressions by using a larger sample including firms from eight 2-digit COD-emitting industries and find qualitatively similar results.[33] Thus, overall, the data provide support for the assumption that pollution control costs depend on agglomeration economies in the county.

8. Conclusion

This paper takes an initial step toward trying to explain why improving environmental outcomes may be difficult in rapidly urbanizing economies. We develop a model to analyze the interaction between agglomeration economies and environmental regulation

[33] The eight major COD-emitting industries are: processing of food from agricultural products (CIC 13), food (14), beverages (15), textiles (17), paper and paper products (22), raw chemical materials and chemical products (26), medicines (27), and smelting and pressing of ferrous metals (32).

in a context where entrepreneurs can decide where to live/locate their firms and over time entry/exit can occur. We find that consideration of agglomeration economies and firm relocation as well as entry/exit could change some of the classic results about the effectiveness of environmental policy. Perhaps the most striking result is that, although an emission standard is effective in reducing aggregate pollution in the short run (when firm location and the number of firms is fixed) for economies that are urbanizing and characterized by partial agglomeration, it could instead exacerbate pollution problems once endogenous firm relocation occurs. In general, endogenous relocation can work to undermine the effectiveness of regulation and even lead to regulations being counter-productive, i.e., worsening environmental conditions by increasing pollution and/or pollution damages. We show that this is possible under both a uniform standard that applies across all regions, as well as a differentiated standard that becomes more stringent in the larger (urban) area. In addition, it can arise even when firms endogenously enter or exit in response to the regulation. The key reason is that environmental regulation that encourages clean production can magnify the impact of agglomeration forces, leading to greater concentrations of firms and pollution that can more than offset the direct effect of the regulation. We suggest that these results could provide at least a partial explanation for persistent environmental problems in urbanizing economies. These economies could experience a period of stagnation or even reversal in environmental quality before they reach the stage where tightening environmental standards will lead to improved environmental quality.

We present initial empirical evidence in the context of water pollution control in China that both demonstrates the importance of agglomeration economies in determining the impacts of environmental regulation and supports our theoretical model specification and predictions. We find that, although China's water quality regulation has a direct negative effect on the total number of firms in the regulated industries, it also enhances the effect of agglomeration economies, which tends to increase the number of firms. This interaction effect offsets the direct effect of the water quality regulation and even outweighs the direct effect of the regulation in counties where agglomeration economies are sufficiently large. Similarly, we show that agglomeration economies work to reduce the effect of regulation on total emissions reduction, i.e., the reduction in emissions due to the regulation is lower in counties with larger agglomeration economies. Again, when agglomeration economies are sufficiently large, the regulation can actually lead to an increase in emissions. We show that this is the case for about one-third of the counties in our sample. Moreover, our estimates suggest that in the aggregate, the regulations governing COD emissions from the pulp and paper industry in China likely led to an overall increase (rather than the intended decrease) in COD emissions (relative to the predicted level in the absence of regulation). This suggests that the forces we identify in our theoretical model can be important in real-world policy contexts as well. Finally, as a possible explanation for our results, we find evidence that COD-emitting firms face lower abatement costs when located in counties with larger agglomeration economies, which provides empirical support for our theoretical model showing that agglomeration benefits are greater with more stringent regulation.

It might be tempting to conclude from our analysis that environmental standards are, or at least can be, “bad” at certain stages of development. However, a perverse environmental result is not necessarily less efficient because the economic benefit of agglomeration may outweigh the environmental cost when more firms agglomerate in one region. As we show in the context of a differentiated standard, when the spatial distribution of firms is not optimal (as is likely, given the pollution and location externalities), increasing the stringency of the standard can increase or decrease aggregate utility (and hence welfare). Moreover, in the presence of both pollution and agglomeration economies, the theory of the second best suggests that multiple instruments are needed to achieve the first-best outcomes. From an environmental perspective, when pollution stems from output production, policies that encourage dispersion of firms and economic activity across space are preferred to policies that encourages agglomeration, which are often prescribed for economic development. To strike a balance between economic development that exploits agglomeration economies and environmental protection, regulators could consider coupling environmental regulation with a regional development policy designed to address relocation externalities to achieve both the optimal spatial distribution of firms and the desired level of emissions from individual firms.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jeem.2022.102754.

References

Ahlfeldt, G.M., Pietrostefani, E., 2019. The economic effects of density: a synthesis. J. Urban Econ. 111, 93–107.
Arnott, R., Hochman, O., Rausser, G.C., 2008. Pollution and land use: optimum and decentralization. J. Urban Econ. 64, 390–407.
Bartik, T.J., 1991. Who Benefits from State and Local Economic Development Policies? Kalamazoo. W.E. Upjohn Institute, MI.
Becker, R., Henderson, V., 2000. Effects of air quality regulations on polluting industries. J. Polit. Econ. 108 (2), 379–421.
Blackman, A., Li, Z., Liu, A.A., 2018. Efficacy of command-and-control and market-based environmental regulation in developing countries. Ann. Rev. Resour. Econ. 10, 381–404.
Borck, R., Pflüger, M., 2019. Green cities? Urbanization, trade, and the environment. J. Reg. Sci. 59, 743–766.
Borck, R., Tabuchi, T., 2019. Pollution and city size: can cities be too small? J. Econ. Geogr. 19, 995–1020.
Brandt, L., Van Biesebroeck, J., Zhang, Y., 2012. Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. J. Dev. Econ. 97, 339–351.
Burkitt, L., Spegele, B., 2013. Why leave job in beijing? To breathe. Wall St. J. April 14.
Cai, X., Lu, Y., Wu, M., Yu, L., 2016. Does environmental regulation drive away inbound foreign direct investment? Evidence from a quasi-natural experiment in China. J. Dev. Econ. 123 (November), 73–85.
Carozzi, F., Roth, S., 2019. Dirty Density: Air Quality and the Density of American Cities. CEP Discussion Papers DP1635, Centre for Economic Performance.

Chernwichan, J., Copeland, B.R., Taylor, M.S., 2017. Trade and the environment: new methods, measurements, and results. Ann. Rev. Resour. Econ. 9, 59–85.
Christmann, P., Taylor, G., 2001. Globalization and the environment: determinants of firm self-regulation in China. J. Int. Bus. Stud. 32 (3), 439–458.
Copeland, B., Taylor, S., 2013. Trade and the Environment: Theory and Evidence. Princeton University Press, Princeton and Oxford.
Cullen, J., 2013. Measuring the environmental benefits of wind generated electricity. Am. Econ. J. Econ. Pol. 5 (4), 107–133.
Dechezlepretre, A., Sato, M., 2017. The impacts of environmental regulations on competitiveness. Rev. Environ. Econ. Pol. 11 (2), 183–206.
Dhiraj, A.B., 2017. Top 20 most polluting rivers in the world. Geoworld Magzine. June 9, 2017. https://ceoworld.biz/2017/06/09/top-20-most-polluting-rivers-in-the-world-chinas-yangtze-and-indias-ganges-tops/.
Duranton, G., Puga, D., 2004. Micro-foundations of urban agglomeration economies. In: Henderson, J.V., Jacques-François, T. (Eds.), Handb. Reg. Urban Econ. 4, 2063–2117.
Elbers, C., Withagen, C., 2004. Environmental policy, population dynamics and agglomeration. Contrib. Econ. Anal. Pol. 3 (2), 1–23.
Ellison, G., Glaeser, E.L., Kerr, W.R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. Am. Econ. Rev. 100 (3), 1195–1213.
Fowlie, M., Reguant, M., Ryan, S., 2016. Market-based emissions regulation and industry dynamics. J. Polit. Econ. 124 (1), 249–302.
Fowlie, M., Muller, N., 2019. Market-based emissions regulation when damages vary across sources: what are the gains from differentiation?". J. Assoc. Environ. Resour. Econ. 6 (3), 593–632.
Fugita, M., Ogawa, H., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Reg. Sci. Urban Econ. 12, 161–196.
Gillingham, K., Rapson, D., Wagner, G., 2016. The rebound effect and energy efficiency policy. Rev. Environ. Econ. Pol. 10 (1), 68–88.
Glaeser, E.L., 1998. Are cities dying? J. Econ. Perspect. 12, 139–160.
Glaeser, E.L., Gottlieb, J.D., 2008. The economics of place-making policies. *Brookings Papers on Economic Activity* 2008 (Spring) 155–253.
Glaeser, E., Kahn, M.E., 2010. The greenness of cities: carbon dioxide emissions and urban development. J. Urban Econ. 67 (3), 404–418.
Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in cities. J. Polit. Econ. 100 (6), 1126–1152.
Goldsmith-Pinkham, P., Sorkin, I., Swift, H., 2020. Bartik instruments: what, when, why, and how. Am. Econ. Rev. 110 (8), 2586–2624.
Grazi, F., van den Bergh, J.C.J.M., van Ommeren, J.N., 2008. An empirical analysis of urban form, transport, and global warming. Energy J. 29 (4), 97–122.
Greenstone, M., 2002. The impacts of environmental regulations on industrial activity: evidence from the 1970 and 1977 clean air act amendments and the census of manufactures. J. Polit. Econ. 110 (6), 1175–1219.
Haskel, Jonathan E., Pereira, Sonia C., Slaughter, Matthew J., 2007. Does inward foreign direct investment boost the productivity of domestic firms? Rev. Econ. Stat. 89 (3), 482–496.
Henderson, J.V., 1977. Externalities in a spatial context: the case of air pollution. J. Publ. Econ. 7 (1), 89–110.
Henderson, J.V., 2003. Marshall's scale economies. J. Urban Econ. 53 (1), 1–28.
Henderson, J.V., 2010. Cities and development. J. Reg. Sci. 50 (1), 515–540.
Hering, L., Poncet, S., 2014. Environmental policy and exports: evidence from Chinese cities. J. Environ. Econ. Manag. 68, 296–318.
Holland, S.P., 2012. Emissions taxes versus intensity standards: second-best environmental policies with incomplete regulation. J. Environ. Econ. Manag. 63 (3), 375–387.
Holland, S.P., Hughes, J.E., Knittel, C.R., 2009. Greenhouse gas reductions under low carbon fuel standards? Am. Econ. J. Econ. Pol. 1 (1), 106–146.
Holland, S., Yates, A., 2015. Optimal trading ratios for pollution permit markets. J. Publ. Econ. (125), 16–27.
Hosoe, M., Naito, T., 2006. Transboundary pollution transmission and regional agglomeration effects. Pap. Reg. Sci. 85 (1), 99–120.
IQAir, 2022. World's Most Polluted Cities (Historical Data 2017-2021). https://www.iqair.com/world-most-polluted-cities?sort=-rank&page=2&perPage=50&cities=. (Accessed 1 October 2022). accessed.
Kahn, M.E., Li, P., Zhao, D., 2015. Water pollution progress at borders: the role of changes in China's political promotion incentives. Am. Econ. J. Econ. Pol. 7 (4), 223–242.
Kahn, M.E., Mansur, E.T., 2013. Do local energy prices and regulation affect the geographic concentration of employment? J. Publ. Econ. 101, 105–114.
Kim, L., 2021. Schools in New Delhi to shut down for a week due to high levels of air pollution. Forbes Business 13, 2021.
Kim, S., 1995. Expansion of markets and the geographic distribution of economic activities: the trends in U. S. Regional manufacturing structure, 1860-1987. Q. J. Econ. 110 (4), 881–908.
Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Econ. 99, 483–499.
Krugman, P., 1998. What's new about the new economic geography? Oxf. Rev. Econ. Pol. 14 (2), 7–17.
Kyriakopoulou, E., Xepapadeas, A., 2011. Spatial location decision under environmental policy and housing externalities. Environ. Econ. Pol. Stud. 13 (3), 195–217.
Kyriakopoulou, E., Xepapadeas, A., 2013. Environmental policy, first nature advantage and the emergence of economic clusters. Reg. Sci. Urban Econ. 43 (1), 101–116.
Kyriakopoulou, E., Xepapadeas, A., 2017. Atmospheric pollution in rapidly growing industrial cities: spatial policies and land use patterns. J. Econ. Geogr. 17, 607–634.
Lange, A., Quaas, M.F., 2007. Economic geography and the effect of environmental pollution on agglomeration. B E J. Econ. Anal. Pol. 7 (1), 1935, 1682.
Latzko, D.A., 2018. The Geographic Concentration of Economic Activity across the Eastern United States, 1820-2010. Working Paper. Pennsylvania State University.
Levinson, A., Taylor, M.S., 2008. Unmasking the pollution haven effect. Int. Econ. Rev. 49, 223–254.
List, J.A., McHone, W.W., Millimet, D.L., 2003. Effects of air quality regulation on the destination choice of relocating plants. Oxf. Econ. Pap. 55 (4), 657–678.
List, J.A., McHone, W.W., Millimet, D.L., 2004. Effects of environmental regulation on foreign and domestic plant births: is there a home field advantage? J. Urban Econ. 56 (2), 303–326.
Lucas, R.E., Rossi-Hansberg, E., 2002. On the internal structure of cities. Econometrica 70, 1445–1476.
Markusen, J., Morey, E.R., Olewiler, N., 1993. Environmental policy when market structure and plant locations are endogenous. J. Environ. Econ. Manag. 24, 69–86.
Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71, 1695–1725.
Mendelsohn, R., 1986. Regulating heterogeneous emissions. J. Environ. Econ. Manag. 13, 301–312.
Moretti, E., 2004. Workers' education, spillovers, and productivity: evidence from plant-level production functions. Am. Econ. Rev. 94 (3), 656–690.
Muller, N.Z., Mendelsohn, R., 2009. Efficient pollution regulation: getting the prices right. Am. Econ. Rev. 99 (5), 1714–1739.
Ottaviano, G.I.P., Takatoshi, T., Thisse, J.-F., 2002. Agglomeration and trade revisited. Int. Econ. Rev. 43, 409–435.
Pargal, S., Wheeler, D., 1996. Informal regulation of industrial pollution in developing countries: evidence from Indonesia. J. Polit. Econ. 104 (6), 1314–1327.
Pflüger, M.P., 2021. City size, pollution and emission policies. J. Urban Econ. 126, 103391.
Porter, M.E., 1990. The Competitive Advantage of Nations. The Free Press, New York.
Regnier, C., Legras, S., 2018. Urban structure and environmental externalities. Environ. Resour. Econ. 70, 31–52.
Rosenthal, S.S., Strange, W.C., 2003. Geography, industrial organization, and agglomeration. Rev. Econ. Stat. 85 (2), 377–393.
Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Jacques-François, T. (Eds.), Handb. Reg. Urban Econ. ume 4, 2119–2171.
Shapiro, J.M., 2006. Smart cities: quality of life, productivity, and the growth effects of human capital. Rev. Econ. Stat. 88 (2), 324–335.
Shapiro, J.S., Walker, R., 2018. Why is pollution from U.S. Manufacturing declining? The roles of environmental regulation, productivity, and trade. Am. Econ. Rev. 108, 3814–3854, 2018.
Tabuchi, T., Thisse, J.-F., 2006. Regional specialization, urban hierarchy, and commuting costs. Int. Econ. Rev. 47 (4), 1468–2354.
Van Marrewijk, C., 2005. Geographical Economics and the Role of Pollution on Location. *Tinbergen Institute Discussion Paper* TI 2005-018/2.
Vance, C., Hedel, R., 2008. On the link between urban form and automobile use: evidence from German survey data. Land Econ. 84 (1), 51–65.
Walker, W.R., 2011. Environmental regulation and labor reallocation: evidence from the clean air act. Am. Econ. Rev. 101, 442–447.
Wang, C., Wu, J., 2011. Locational amenities, increasing returns, and urban development. J. Econ. Geogr. 11, 687–707.

Wang, C., Wu, J., Zhang, B., 2018. Environmental regulation, emissions and productivity: evidence from Chinese COD-emitting manufacturers. J. Environ. Econ. Manag. 92, 54–73.
World Bank, 2018. World Development Indicators. The World Bank, Washington DC. https://datacatalog.worldbank.org/dataset/world-development-indicators. (Accessed 26 July 2018). accessed.
Wu, J., 2019. Agglomeration: economic and environmental impacts. Ann. Rev. Resour. Econ. 11 (October), 419–438.
Wu, J., Reimer, J., 2016. Pollution mobility, productivity disparity, and the spatial distribution of polluting and nonpolluting firms. J. Reg. Sci. 56, 615–634.
Wu, J., Segerson, K., Langpap, C., 2022. Spatial dispersion vs. Agglomeration: a new explanation of the environmental kuznets curve. Working paper.
Zeng, D., Zhao, L., 2009. Pollution havens and industrial agglomeration. J. Environ. Econ. Manag. 58 (2), 141–153.
Zheng, S., Kahn, M.E., 2013. Understanding China's urban pollution dynamics. J. Econ. Lit. 51 (3), 731–772.
Zhou, R., Bi, X., Segerson, K., 2017. Evaluating voluntary environmental programs with spillover effects. J. Assoc. Environ. Resour. Econ. 7 (1), 145–180.