Growth and the pollution convergence hypothesis: a nonparametric approach.

C. Ordás Criado* S. Valente† T. Stengos‡

September 17, 2009

Abstract

The pollution-convergence hypothesis is formalized in a neoclassical growth model with optimal emissions reduction: pollution growth rates are positively correlated with output growth (scale effect) but negatively correlated with emission levels (defensive effect). This dynamic law is empirically tested for two major and regulated air pollutants - nitrogen oxides (NOX) and sulfur oxides (SOX) - with a panel of 25 European countries spanning over years 1980-2005. Traditional parametric models are rejected by the data. However, more flexible regression techniques - semiparametric additive specifications and fully nonparametric regressions with discrete and continuous factors - confirm the existence of the predicted positive and defensive effects. By analyzing the spatial distributions of per capita emissions, we also show that cross-country pollution gaps have decreased over the period for both pollutants and within the Eastern as well as the Western European areas. A Markov modeling approach predicts further cross-country absolute convergence, in particular for SOX. The latter results hold in the presence of spatial non-convergence in per capita income levels within both regions.

JEL Classification numbers: C14, C23, Q53

Key Words: Air pollution, convergence, economic growth, mixed nonparametric regressions, distribution dynamics.

---

*Corresponding author: cordas@ethz.ch, Center for Energy Policy and Economics (CEPE), ETH Zurich, Switzerland.

†Center of Economic Research (CER), ETH Zürich, Switzerland.

‡Department of Economics, University of Guelph, N1G2W1, Ontario, Canada.

1 Introduction

The continued expansion of output and the intensive use of natural resources have been a cornerstone of economic prosperity. The other side of the coin is the increasing pressure exerted on the environment by the growth process. This conflictive dynamic of economic growth calls for the use of more efficient technologies as well as defensive expenditures to curve down pollution and alleviate health or environmental damages. From the normative side, it raises the question of what policy measures should be applied to achieve sustainable growth, *i.e* "a balanced growth path with increasing environmental quality and ongoing growth in income per capita", see Brock and Taylor (2004a, p.2-3). From a positive point of view, the ability of market forces to freely balance the pros and cons of an extensive development is called into question by natural scientists. Capturing the underlying dynamic of the pollution-GDP relationship with data is essential from both perspectives. This paper focuses on the per capita air pollutants' emissions and GDP dynamic and falls within the scope of both approaches. It provides flexible estimates of the relationship by considering level as well as growth variables within a simple specification derived from a growth model with optimal pollution control.

Economic analysis has developed numerous approaches to tackle the GDP-pollution relationship and to explain the most prominent stylized facts. The most controversial empirical finding, known as the 'Environmental Kuznets Curve (EKC)', states that a U-inverted relationship exists between GDP and some specific pollutants. This particular shape has been replicated in a variety of theoretical frameworks[1] and reflects a pattern that is compatible with sustainable growth. More generally, a key requirement for settling the trade-off between economic growth and environmental quality in the long run is convergence in pollution levels - that is, achieving a path along which harmful emissions and the associated damage are declining, or at least bounded and stabilized, from some instant onwards. The crucial instruments to fulfill this objective are abatement activities and emission-reducing technical progress - *i.e.* a dynamic process whereby the economy is able to decrease the emission intensity of production over time. At the theoretical level, the pollution-convergence hypothesis has been studied in Brock and Taylor (2004b) by means of a 'Green Solow Model' - *i.e.* a neoclassical growth model in which the saving rate and the propensity to spend in abatement are exogenously fixed. They show that economic growth and increasing environmental quality require a sufficiently high rate of emission-reducing technical progress, and that economic growth is actually exploited to generate positive feedback effects on environmental quality through investment in abatement activities.

Building on this point, we analyze a possible micro-foundation of the convergence hypothesis by endogeneizing the propensities to consume and to invest in clean technologies in a neoclassical growth model à la Ramsey. Specifically, we

[1] See Brock and Taylor (2004b), Andreoni and Levinson (2001) or Stockey (1998) among others.

introduce capital accumulation, emission-reducing and labor-augmenting technological progress in a model of optimal emission reduction through investment in clean technologies, see Van der Ploeg and Withagen (1991). We show that, along the optimal path, the instantaneous growth rate of emissions per capita is (i) negatively correlated with the level of emissions per capita (defensive effect) and (ii) positively correlated with the growth rate of output per capita (scale effect). The model sets a simple equation that allows to test the presence of both forces at the empirical level under the assumption of optimal control for pollution. Moreover, since the propensity to save and the share of investment devoted to reduce the emission intensity are both endogenous, this framework provides an alternative micro-foundation of the pollution-convergence hypothesis with respect to Brock and Taylor (2004b).

The existence of both a scale and a defensive effect is tested empirically for several air pollutants' emissions across a panel of 25 Eastern and Western European countries over the period 1980-2005. Our test-equation has a very similar structure to the β-type formulation used in the empirical growth literature. We base the empirical evidence on a large set of regression methods. Firstly, we show that the standard parametric β-type regressions often used to investigate pollution convergence across countries is likely to be misspecified. Secondly, a flexible regression approach is proposed to address this issue : we compute semiparametric as well as multivariate nonparametric regressions that better captures nonlinearities and interactions between the regressors. The results appear to be consistent with our theoretical model. More precisely, we find that more economic activity generates more pollution but emissions-reducing technological progress and/or increasing investment in clean technologies act against the scale effect. We further show with a distribution analysis that the latter results are obtained in the context of decreasing per capita pollution gaps between countries and increasing/stable per capita income disparities.

The structure of the paper is as follows. The next section describes our theoretical model. Section 3 presents the empirical methodology, starting with the parametric and the nonparametric regression approaches in sections 3.1 and 3.2 and proceeding with the distribution dynamics analysis in section 3.3. The data and results are shown in sections 4 and 5. We conclude in section 6.

2 A simple model of growth and optimal emission reduction

The relationship between economic growth and pollution dynamics has been investigated by an important, and still growing, body of theoretical literature. The standard approach, initiated by Keeler et al. (1971), focuses on the interactions between capital accumulation and emission intensities. The link between pollution levels and economic activity is represented by an emission function, in which the production process generates emissions, but economic growth may induce positive feedback effects: if the economy invests resources in the development of cleaner technologies, the elasticity of emissions to output can be reduced over time. In this framework, Van der Ploeg and Withagen (1991) provide a comprehensive treatment of capital-pollution dynamics in the Ramsey-Cass-Koopmans model, and Bovenberg and Smulders (1995) and Stockey (1998) extend the analysis to incorporate endogenous growth theories.

The hypothesis of convergence in output and pollution levels, which is the central issue of our paper, is formally addressed in Brock and Taylor (2004b) by means of a 'Green Solow Model' - i.e. a neoclassical growth model with labor-augmenting technical progress, in which the saving rate and the propensity to spend in abatement are exogenously fixed. In this section, we provide a simple micro-foundation of the convergence hypothesis by endogeneizing the propensities to consume and to invest in clean technologies in a neoclassical growth model *à la* Ramsey. This allows us to derive the basic equation to be employed in the empirical analysis from the optimality conditions of a centralized social problem : in our model, the pollution-income relation is determined by the saddle path describing the transitional dynamics of emission levels.[2] The model presented may be considered as an extension of the "clean-technologies variant" of Van der Ploeg and Withagen (1991, sect.8), which includes capital accumulation, emission-reducing and labor-augmenting technological progress.[3]

2.1 The Ramsey setting

As our empirical analysis will focus on the dynamics of emissions per capita, we will treat pollution as a flow-variable that affects private utility in per capita terms. Time is continuous and indexed by $t \in [0, \infty)$, and the model economy is characterized by the following assumptions:[4]

[2]Our micro-founded model is alternative to, but not conflicting with the Solow-type model of Brock and Taylor (2004b). These authors acknowledge the importance of endogenously determined abatement efforts, and briefly assess its potential implications for the time path of pollution per capita (Brock and Taylor 2004b, pp. 48-50).

[3]Van der Ploeg and Withagen (1991) study pollution and welfare-reducing emissions in different variants of the Ramsey-Cass-Koopmans model, with and without capital accumulation. Among these different models, the closest to our analysis is the "clean-technologies variant" (Van der Ploeg and Withagen 1991, sect.8) which however abstracts from capital accumulation as well as technological progress. The present model can thus be interpreted either as an extension of the "clean-technologies model" to include capital accumulation and technical progress, or as an extension of the Ramsey model with flow-pollution and capital ((Van der Ploeg and Withagen 1991, sect.3)) to include optimal investment in emission reduction.

[4]Using standard notation, we define $\dot{Z} \equiv dZ/dt$ as the time-derivative of the generic variable $Z(t)$, and $G_Q \equiv \partial G/\partial Q$ and $G_{QQ} \equiv \partial^2 G/\partial Q^2$ as the partial derivatives of the generic function $G(Q)$.

$$Y(t) = F(K(t), B(t) N(t)), \quad B(t) = B_0 e^{\pi t}, \quad N(t) = N_0 e^{nt}, \tag{1}$$
$$\dot{K}(t) = Y(t) - C(t) - X(t) - \delta K(t), \tag{2}$$
$$P(t) = \tau(t) Y(t), \tag{3}$$
$$U(t) = U(\bar{c}(t), \bar{p}(t)), \quad U_{\bar{c}} > 0, \quad U_{\bar{c}\bar{c}} \leq 0, \quad U_{\bar{p}} < 0, \quad U_{\bar{p}\bar{p}} < 0, \tag{4}$$

Technology (1) assumes that aggregate output, Y, is produced by means of capital, K, and efficient labor, BN, according to a linearly homogeneous production function $F(K, BN)$ displaying positive and strictly decreasing marginal productivities in both factors, and satisfying the Inada conditions. Population N grows at the exogenous rate $n > 0$, and labor efficiency B grows at the given rate of labor-augmenting technical progress $\pi > 0$. Expression (2) is the accumulation constraint, where $\delta \geq 0$ is the rate of physical depreciation of capital: net investment equals output minus the sum of the economy's expenditures, represented by consumption, C, and defensive expenditures, X. By defensive expenditures we mean resources devoted to activities that reduce the emission intensity of the production sector, and we will label it as 'ICT' (investment in cleaner technology). The pollution function (3) asserts that aggregate emissions per unit of time, P, are proportional to aggregate output, and τ represents the aggregate emission intensity. Expression (4) defines private utility, U, as a function of consumption per capita ($\bar{c} \equiv C/N$) and emissions per capita ($\bar{p} \equiv P/N$), where $U_{\bar{p}} < 0$ and $U_{\bar{p}\bar{p}} < 0$ guarantee that the disutility from pollution is convex - i.e. the marginal health damage increases more than proportionally to emissions per capita.

In order to obtain a full analytical characterization of optimal dynamics, we model (i) the process of emission reduction, and (ii) the trade-off between consumption and health damage, by means of two specifications often exploited in related literature. In the first regard, we follow Brock and Taylor (2004b), and assume that the aggregate emission intensity is given by

$$\tau(t) \equiv \Omega(t) \cdot \left[1 - \frac{X(t)}{Y(t)}\right]^{\varepsilon}, \quad \varepsilon > 1, \tag{5}$$

where $\Omega(t)$ is the *baseline emission intensity*, and the second term is a function representing the effects of ICT. The share of output devoted to defensive expenditures, X/Y, will be called *ICT effort*, and corresponds to the propensity to invest in cleaner technologies, bounded between zero and unity. From (5), maximal ICT effort ($X = Y$) implies zero emissions, whereas zero defensive expenditures imply that the emission intensity equals the baseline level. Also the baseline intensity varies over time, as it is affected by technological progress. As shown by Brock and Taylor (2004b), along balanced growth paths, unbounded increases in pollution per capita can be avoided only if the rate of emission-reducing progress - i.e. the effects of technical improvements that reduce $\Omega(t)$ over time - is at least equal to the rate of output-augmenting technical progress. In the present model, this sustainability condition corresponds to $\dot{\Omega}(t)/\Omega(t) \leq -\pi$ (see Lemma 3 in Appendix). In order to ensure that it is technically feasible to obtain stationary pollution per capita in the long run, we assume symmetric rates of emission-reducing and labor-augmenting technical progress, and set $\Omega(t) = \Omega_0 e^{-\pi t}$. Alternative assumptions that satisfy the sustainability condition (i.e. $\Omega(t) = \Omega_0 e^{-\omega t}$ with $\omega > \pi$) would complicate the analysis of steady-state equilibria without affecting the main results concerning the convergence hypothesis.[5]

The second assumption is related to private preferences. Since the optimal control problem comprises two interacting control variables (consumption and defensive expenditures), we will restrict our attention to an instantaneous utility function that allows us obtain an analytical characterization of long-run dynamics, which is the central aim of this section. In this regard, a convenient specification is

$$U(\bar{c}(t), \bar{p}(t)) = \sigma \ln \bar{c}(t) - \varsigma \bar{p}(t)^{\theta}, \quad \theta > 1, \tag{6}$$

where $\sigma > 0$ and $\varsigma > 0$ are the weights on utility from consumption and disutility from pollution, respectively. Function (6) satisfies all the properties listed in (4), consistently with the conditions for a well-behaved problem with neoclassical production functions : $\theta > 1$ ensures increasing marginal damage (see (Van der Ploeg and Withagen 1991)).

The optimal path is defined as a sequence of consumption levels and defensive expenditures which maximizes the present value of the discounted stream of utilities

$$\int_0^{\infty} U(\bar{c}(t), \bar{p}(t)) e^{-\rho t} dt, \tag{7}$$

subject to the accumulation constraint (2), the pollution function (3), and the non-negativity constraint $K(t) \geq 0$ in each t, for a given initial stock $K(0) = K_0$ and given parameters B_0, N_0, Ω_0. This problem can be solved more easily by denoting ICT effort as

$$\chi(t) \equiv X(t)/Y(t),$$

and normalizing the relevant variables in terms of labor-efficiency units. Setting $y \equiv Y/(BN)$ and $k \equiv K/(BN)$, the homogeneous production function in (1) yields the intensive form $y = f(k) = F(k, 1)$, where f_k coincides with the marginal product of capital. As a consequence, equations (2)-(3) can be written as

$$\dot{k}(t) = f(k(t))[1 - \chi(t)] - c(t) - (\delta + n + \pi) k(t), \tag{8}$$

$$p(t) = \Omega(t)[1 - \chi(t)]^{\varepsilon} f(k(t)), \tag{9}$$

[5]Assuming more intense emission-reducing technical progress (i.e. $\dot{\Omega}/\Omega < -\pi$) would not affect the main results: pollution per capita would tend to a peculiar steady-state level (zero) in the long run, confirming the convergence hypothesis. We assume perfectly symmetric rates of emission-reducing and labor-augmenting technical progress because asymmetric rates would imply additional technical difficulties without any gain for the present analysis. This point is clarified in the Appendix - see the discussion below Lemma 3.

where $c \equiv C/(BN)$ and $p \equiv P/(BN)$ are 'normalized' consumption and pollution, respectively. Since the arguments in the utility function respectively equal $\bar{c} = cB$ and $\bar{p} = pB$, the optimal path can be found by maximizing (7) subject to (8)-(9), using the sequences of $c(t)$ and $\chi(t)$ as control variables. As shown in the Appendix, the necessary conditions for optimality in an interior solution yield

$$\dot{c}(t)/c(t) = \left[f_k(t)(1-\chi(t))\left(1-\varepsilon^{-1}\right)\right] - (\rho+\delta+n+\pi), \tag{10}$$

$$1-\chi(t) = \Gamma f(k(t))^{-\frac{\theta-1}{\varepsilon\theta-1}} c(t)^{-\frac{1}{\varepsilon\theta-1}}, \tag{11}$$

where Γ is an exogenous constant, and $f_k(t) \equiv f_k(k(t))$. Expression (10) is the growth rate of normalized consumption along the optimal path, which is different from the usual Keynes-Ramsey rule due to the presence of abatement effort, $\chi(t)$, and of the elasticity factor $1-\varepsilon^{-1}$ that quantifies the distortion in the marginal benefit from accumulation induced by welfare-reducing emissions.[6] Condition (11) determines the optimal propensity to spend in abatement, which exhibits a precise link with the time-paths of $k(t)$ and $c(t)$: if consumption and output increase (decrease) along the optimal path, the abatement effort $\chi(t)$ increases (decreases) as well. The dynamic properties of the optimal path can be analyzed as follows. Omitting time-arguments for simplicity, define the right hand side of (11) as a function

$$\Phi(k,c) \equiv \Gamma f(k)^{-\frac{\theta-1}{\varepsilon\theta-1}} c^{-\frac{1}{\varepsilon\theta-1}}, \tag{12}$$

which implies $\Phi_k < 0$ and $\Phi_c < 0$. Substituting $1-\chi = \Phi(k,c)$ in (8) and (10), the resulting differential system is autonomous, and may be written as

$$\dot{k} = f(k)\Phi(k,c) - c - (\tilde{\rho}-\rho)k, \tag{13}$$

$$\dot{c} = f_k(k)\Phi(k,c)\left(1-\varepsilon^{-1}\right)c - \tilde{\rho}c, \tag{14}$$

where we have defined $\tilde{\rho} \equiv \rho+\delta+n+\pi$. Let us denote by (c^{ss}, k^{ss}) the couple of values representing the simultaneous steady-state equilibrium of system (13)-(14). As shown in the Appendix, the steady state exists and is unique for any well-behaved neoclassical production function. Also the stability properties of the steady state are quite general, and do not require assuming specific technologies:

Lemma 1 *The simultaneous steady-state equilibrium (c^{ss}, k^{ss}) of system (13)-(14) is saddle-point stable. Given the initial condition $k(0) = k_0$, the optimal path is unique and implies convergence towards (c^{ss}, k^{ss}).*

[6] In general, the Keynes-Ramsey rule asserts that the sign of consumption growth rates is determined by the difference between the marginal benefit from capital accumulation (i.e. the opportunity cost of postponing consumption) and the marginal benefit from current consumption. In the neoclassical model without pollution, the marginal benefit from accumulation is simply the marginal product of capital net of depreciation. In the current model, instead, the marginal benefit from accumulation is represented by the term in square brackets in (10), which is lower than f_k. The reason is that the optimal path is chosen taking into account that higher capital implies *ceteris paribus* higher pollution, so that there is a wedge between capital profitability and private benefits from accumulation.

Lemma 1 has three main implications. First, convergence towards (c^{ss}, k^{ss}) implies that the propensity to spend in clean technologies and the marginal product of capital are constant in the long run. The asymptotic value of ICT effort, χ^{ss}, is determined by condition (11), whereas the long-run level of capital profitability is determined by

$$f_k^{ss} = \frac{\rho + \delta + n + \pi}{(1 - \chi^{ss})(1 - \varepsilon^{-1})}, \tag{15}$$

which follows from imposing stationarity in (10). Expression (15) implies that normalized capital k^{ss} will be lower than in the Ramsey model - where the modified golden rule $f_k^{ss} = \rho + \delta + n + \pi$ holds.[7] The second implication of Lemma 1 is that the economy displays balanced growth in the long run. Since $c = C/NB$ and $k = K/NB$, as well as ICT effort $\chi = X/Y$, achieve stationary values, the aggregate output, capital, and expenditures grow asymptotically at the balanced rate

$$\lim_{t\to\infty} \frac{\dot{Y}(t)}{Y(t)} = \lim_{t\to\infty} \frac{\dot{K}(t)}{K(t)} = \lim_{t\to\infty} \frac{\dot{C}(t)}{C(t)} = \lim_{t\to\infty} \frac{\dot{X}(t)}{X(t)} = n + \pi, \tag{16}$$

which implies that per capita output and expenditure levels grow at the rate of labor-augmenting technical progress, π.

The third implication of Lemma 1 is that pollution per capita, $\bar{p}(t) = p(t)B(t)$, converges to a constant steady-state level. From (3) and (1), the dynamics of $\bar{p}(t)$ are governed by

$$\bar{p}(t) = \Omega(t)\,\Phi(k(t), c(t))^{\varepsilon} f(k(t))\,B(t) = \Omega_0 B_0 \Gamma^{\varepsilon} f(k(t))^{\frac{\varepsilon-1}{\varepsilon\theta-1}} c(t)^{-\frac{\varepsilon}{\varepsilon\theta-1}}, \tag{17}$$

which implies

$$\lim_{t\to\infty} \bar{p}(t) = \bar{p}^{ss} \equiv \Omega_0 B_0 \Gamma^{\varepsilon} f(k^{ss})^{\frac{\varepsilon-1}{\varepsilon\theta-1}} (c^{ss})^{-\frac{\varepsilon}{\varepsilon\theta-1}}. \tag{18}$$

The transitional dynamics of pollution per capita can be studied by re-introducing $\bar{p}(t)$ in the dynamic system (13)-(14). Indeed, pollution per capita is a jump variable like consumption, and it is generally possible to substitute the optimality relations between $c(t)$ and $\bar{p}(t)$ in (14) in order to analyze the joint dynamics of $(k(t), \bar{p}(t))$. Since we have shown that the dynamics of $(k(t), c(t))$ are saddle-point stable, the same dynamic behavior is expected to arise in the $(k(t), \bar{p}(t))$ plane. This result is formally proved below. In order to obtain an explicit relation between pollution per capita and the other endogenous variables of empirical interest, the following analysis assumes that the aggregate technology is Cobb-Douglas, which considerably simplifies the derivations.

[7] This result is in line with different stationary equilibria analyzed in Van der Ploeg and Withagen (1991). Since the planner takes into account the fact that higher capital implies *ceteris paribus* higher pollution, it is optimal to accumulate less capital with respect to the modified golden rule.

2.2 Transitional dynamics of pollution per capita

Suppose that technology (1) takes the Cobb-Douglas form $Y = K^{\alpha} (BN)^{1-\alpha}$, and write normalized output $f(k) = k^{\alpha}$, where $\alpha \in (0,1)$. In this case, the dynamics of pollution per capita and normalized capital are governed by the non-linear system (see Appendix)

$$\begin{aligned} g(\bar{p}(t)) &= \varphi_0 - \varphi_1 k(t)^{\alpha-1-\alpha\frac{1}{\varepsilon}} \bar{p}(t)^{-\left(\theta-\frac{1}{\varepsilon}\right)}, && (19) \\ g(k(t)) &= \varphi_2 k(t)^{\alpha-1-\alpha\frac{1}{\varepsilon}} \bar{p}(t)^{\frac{1}{\varepsilon}} \left(1-\varphi_3 \bar{p}(t)^{-\theta}\right) - \varphi_4, && (20) \end{aligned}$$

where $g(\bar{p}(t)) \equiv (d\bar{p}(t)/dt)/\bar{p}(t)$ and $g(k(t)) \equiv \dot{k}(t)/k(t)$ are the instantaneous growth rates and $(\varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4)$ are exogenous constants, all strictly positive. From (19)-(20), we have the stationary loci

$$\begin{aligned} g(\bar{p}(t)) &= 0 \quad \rightarrow \quad k(t) = \left[\frac{\psi_1}{\psi_0} \bar{p}(t)^{-\left(\theta-\frac{1}{\varepsilon}\right)}\right]^{\frac{1}{1-\alpha+\alpha\frac{1}{\varepsilon}}}, && (21) \\ g(k(t)) &= 0 \quad \rightarrow \quad k(t) = \left[\frac{\psi_2}{\psi_4} \bar{p}(t)^{\frac{1}{\varepsilon}} \left(1-\psi_3 \bar{p}(t)^{-\theta}\right)\right]^{\frac{1}{1-\alpha+\alpha\frac{1}{\varepsilon}}}, && (22) \end{aligned}$$

where locus (21) is strictly decreasing, while (22) is strictly decreasing, in the $(k(t), \bar{p}(t))$ plane. Since (19) and (20) respectively imply $\partial g(\bar{p})/\partial\bar{p} > 0$ and $\partial g(k)/\partial k < 0$, we obtain the phase diagram reported in Figure 1, graph (a): the simultaneous steady-state $(k^{ss}, \bar{p}^{ss})$ is saddle-point stable, with a strictly decreasing stable path. Since $k(t)$ converges to k^{ss} in the long run, given an initial stock $k(0) = k_0$, the associated initial level of pollution per capita $\bar{p}(0)$ is that determined by the saddle path. In particular, if we want to reproduce the transitional dynamics of an economy exhibiting a positive growth rate, we have to assume $k_0 < k_{ss}$, i.e. that the economy accumulates capital at a positive rate during the transition. Given $k_0 < k^{ss}$, the strictly-decreasing saddle path implies that the initial level of pollution per capita is above the long-run value, $\bar{p}(0) > \bar{p}^{ss}$. Hence, the transitional dynamics are characterized by a *decreasing time path of pollution per capita*, as shown in Figure 1, graph (b).

All the above results can be formally obtained by means of a linearization of system (19)-(20), which implies (see Appendix)

$$\begin{aligned} g(\bar{p}(t)) &\approx m_1 (\bar{p}(t) - \bar{p}_{ss}) + m_2 (k(t) - k_{ss}), && (23) \\ g(k(t)) &\approx m_3 (\bar{p}(t) - \bar{p}_{ss}) + m_4 (k(t) - k_{ss}), && (24) \end{aligned}$$

where the coefficients are $m_1, m_2, m_3 > 0$ and $m_4 < 0$. The Jacobian matrix associated with (23)-(24) confirms saddle-point stability and, in particular, yields the equation of the stable arm (see Appendix)

$$(k(t) - k_{ss}) = \phi(\bar{p}(t) - \bar{p}_{ss}), \qquad \phi < 0, \tag{25}$$

where $\phi < 0$ implies a negatively-sloped saddle path. The stable-arm equation allows us to obtain an explicit relation between pollution per capita growth and the other endogenous variables of empirical interest. In fact, substituting (25) in (23), and using (24) to eliminate normalized capital $k(t)$ from the resulting expression, we obtain (see Appendix)

$$g(\bar{p}(t)) \approx \frac{m_1}{\alpha(m_3 + m_4\phi)}(g(\bar{y}(t)) - \pi) + \phi m_2(\bar{p}(t) - \bar{p}_{ss}), \tag{26}$$

where $g(\bar{y}(t))$ is the growth rate of output per capita $\bar{y}(t)$. Collecting the constant terms and checking the signs of the exogenous parameters appearing in (26), we obtain the following result:

Proposition 2 *Along the optimal path, the instantaneous growth rate of emissions per capita is (i) positively related with the growth rate of output per capita and (ii) negatively related with the level of emissions per capita:*

$$g(\bar{p}(t)) \approx H_0 + H_1 g(\bar{y}(t)) - H_2 \bar{p}(t), \tag{27}$$

where $H_1 > 0$ and $H_2 > 0$.

Proposition 2 provides a micro-foundation to the pollution-convergence hypothesis, according to which pollution levels approach steady-state levels determined by the technological and preference parameters of growing economies. The difference with respect to the reduced forms of Solow-type models employed, e.g., in Brock and Taylor (2004b) and Bulte, List and Strazicich (2007), is that these models rule out the optimization of saving rates and of the propensity to in clean technologies. In the present model, instead, relation (27) is directly obtained from the saddle path followed by pollution per capita during the transition, which incorporates all the optimality conditions governing consumption and investment decisions.

Equation (27) suggests a simple way to test the pollution-convergence hypothesis at the empirical level. Consider a discrete-time equation of the type

$$GP_t = \gamma_0 - \beta \ln P_{t-T} + \gamma_1 GY_t, \tag{28}$$

where (GP_t, GY_t, P_{t-T}) represent $(g(\bar{p}(t)), g(\bar{y}(t)), \bar{p}(t))$ in a discrete-time setting with T-year periods growth rates. From (28), the pollution-convergence hypothesis can be verified by testing empirically that β and γ_1 are strictly positive. In view of the fact that (28) is a first-order approximation, it is also recommendable to include an additional regressor which captures the high order effects cleaned out by the Taylor expansion underlying system (23)-(24). Since the deviations arising between the exact non-linear saddle path in Figure 1 and the linearized stable arm (25) are essentially due to the dynamics of capital, the natural hypothesis is to include past levels of output per capita as an additional explanatory variable in (28). Therefore, the convergence model proposed in the following section is given by

$$GP_t = \gamma_0 - \beta \ln P_{t-T} + \gamma_1 GY_t + \gamma_2 \ln Y_{t-T}, \tag{29}$$

where we do not postulate a priori a definite sign for the coefficient γ_2 associated with output levels.

3 Empirical methodology

This section proposes three empirical methodologies which explore the pollution convergence phenomenon captured in the test-equation (29). Firstly, we present in section 3.1 a standard parametric approach often used in growth regressions, along with a specification test that formally checks whether the parametric setting consistently fits the data. Secondly, we introduce more flexible regression methods in section 3.2 as alternatives in case of misspecification under the parametric model. Finally, the distributional analysis described in section 3.3 investigates the dynamics of the cross-sectional gaps in both income and pollution.

3.1 Parametric regressions

As it is common in recent papers, the theoretical model is investigated with panel regressions. Rather than making use of yearly growth rates for pollution and income, our panel estimates are based on four 5-year ($T = 5$) periods, starting in year 1980. As pointed out by Barro and Sala-i Martin (2004, Ch. 11.10), taking shorter periods carries the risk of missing long run adjustments. More precisely, short run growth rates tend to capture short term adjustments around the trend rather than long run convergence. In the presence of business cycles, this leads to an upward bias of the estimates of the convergence speed. Our specification also allows homogeneous groupings of countries to display structural dissimilarities through a group-specific dichotomous variable. Time dummies are also included to account for potential structural breaks and to capture time-specific effects in the relationship. Therefore the panel model is given by

$$GP_{i,t} = \alpha_1 + \alpha_2 D_i + \alpha_{3,t} D_t + \beta log P_{i,t-T} + \delta_1 Y_{i,t-T} + \delta_2 GY_{i,t} + \varepsilon_{i,t} \qquad (30)$$

where:

$GP_{i,t}$: is the growth rate of emissions per capita in the i-th country, measured by the average log changes $(1/T)\log(P_{i,t}/P_{i,t-T})$ over the time span $t-T$ to t;

D_i : is a dummy equal to 1 if the i-th country is an EU15 member and equal to 0 if not;

D_t : are dummy variables for each period t of the panel;

$P_{i,t-T}$: is the (log of the) level of emissions per capita (tons/capita) in the i-th country at time $t-T$;

$GY_{i,t}$: is the growth rate of GDP per capita in the i-th country, measured by the average log changes $(1/T)\log(Y_{i,t}/Y_{i,t-T})$ over the time span $t-T$ to t;

$Y_{i,t-T}$: is the (log of the) level of GDP per capita (in 1990 International Geary-Khamis dollars) in the i-th country at time $t-T$;

$\varepsilon_{i,t}$: is an iid error term.

Indeed, when T is set to the entire length of the time dimension, specification (30) becomes a cross-sectional model where the dynamic component is captured by growth rates over the whole period. Therefore, it can be naturally estimated either with cross-sectional or panel regressions. The latter framework has the advantage of better capturing unobserved heterogeneity and nonlinearities in the relationship.

Given the potential feedback effect of pollution on GDP, regression's coefficients can suffer from endogeneity bias. We address this issue by providing Instrumental Variables (IV henceforth) estimates. Following Barro and Sala-i Martin (2004, Ch12.2.2), we keep the first 5 years of observations out of the sample to build instruments. Therefore, IV versions of the panel regressions are also proposed. We retain $Y_{i,t-1}$ as instrument for $Y_{i,t}$ and $GY_{i,(t-1)-T}$ for $GY_{i,t-T}$.

Finally, two fundamental hypotheses are tested regarding the OLS fits. The null of homoscedasticity is checked with a robust version's heteroscedasticity test[8]. We also apply the specification test by Hsiao, Li and Racine (2007) to check if the parametric linear models provide consistent estimates. More exactly, if $E(y_i, x_i)$ is the true but unknown conditional mean that is approximated by some parametric model $E(y_i, x_i; \varphi)$, this test contrasts the following two hypotheses: $H_0 : E(y_i, x_i) = E(y_i, x_i; \varphi)$ *vs.* $H_1 : E(y_i, x_i) \neq E(y_i, x_i; \varphi)$, almost everywhere. If H_0 is not accepted, more flexible specifications can be explored. This paper considers two alternatives to the linear parametric model (30) : a semiparametric additive model which gives full flexibility to the continuous explanatory components as well as a fully nonparametric regression which allows all kind of interactions between the independent variables, and in particular between the continuous regressors and the group-specific and year-specific dummies.

3.2 Nonparametric regressions

We further proceed to estimate a more flexible version of equation (30) whereby we relax certain functional restrictions that allow for some of the variables to enter parametrically and the others nonparametrically but with a separable structure. This is the partially linear (PLR) additively separable regression model which can be written as

$$GP_{i,t} = \alpha_1 + \alpha_2 D_i + \alpha_{3,t} D_t + \sum_{j=1}^{3} f_j(x_j^c) + \varepsilon_{i,t} \tag{31}$$

where $f_j(x_j^c)$ are three unknown nonlinear functions, one for each jth continuous factors from model (30), *i.e.* $x_1^c = P_{i,t-T}, x_2^c = Y_{i,t-T}, x_3^c = GY_{i,t}$. The first three terms in specification (31) constitute the linear part of the PLR model while the

[8]Greene (2003, Ch.11.4.3)

last term $\sum_{j=1}^{3} f_j(x_j^c)$ is the additive nonparametric component. Compared to the parametric model (30), the PLR setting imposes no restriction on the flexibility of the additive nonparametric factors and it allows a straightforward graphical representation along all its dimensions. The additive block is a quite restrictive special case of the general smooth function $f(x_1^c, x_2^c, x_3^c)$ but it can be estimated more efficiently than a fully nonparametric setting when it represents the true relationship. Fan, Haerdle and Mammen (1998) and Fan and Li (2004) use marginal integration to estimate the components of the additive semiparametric PLR model in equation (31). From another perspective, PLR specifications can also be fitted with generalized additive models's techniques[9]. Wood (2000, 2006) proposes to decompose the flexible additive components in a finite sum of spline terms and to apply penalized least squares combined with a cross-validation to control for the smoothness of the functions. More precisely, let each $f_j(.)$ component be represented by $f_j(z_j) = \sum_{k=1}^{K} \beta_k g_k(.)$, where $g_k(.)$ is a family of K spline basis functions, and let the penalizing roughness term be $\int [f''(z_j)]^2 dz_j$, therefore equation (31) can be estimated by minimizing the following expression:

$$\min_{\alpha_r, \beta_k, \lambda_j} \sum_i \left(y_i - \sum_{r=1}^{R} \alpha_r x_{r,i} - \sum_{j=1}^{J} f_j(z_{j,i}) \right)^2 + \sum_{j=1}^{J} \lambda_j \int f_j''(z_j)^2 dz_j$$

where the $f_j(.)$ terms are replaced by their spline decomposition and the λ_js are determined by generalized cross-validation. In this paper, specification (31) is explored with penalized smoothing splines and its results are reported graphically for each estimated function $\widehat{f_j}(x_j^c)$, with $j = 1, 2, 3.$[10]

The functional restrictions in the parametric model (30) and the nonparametric additive hypothesis of the PLR equation (31) are fully relaxed by estimating

$$GP_{i,t} = f(\mathbf{x}^d, \mathbf{x}^c) + \epsilon_{i,t} \tag{32}$$

where $\mathbf{x}^d = [D_i, \tilde{D}_t]$ are the usual discrete regressors but with D_t defined as a single discrete trend factor and $\mathbf{x}^c = [P_{i,t-T}, Y_{i,t-T}, GY_{i,t}]$ are the continuous explanatory factors. Racine and Li (2004) have recently proposed a new kernel method to estimate nonparametric regressions with mixed independent variables which is consistent in panels with a small time dimension t relative to the individual dimension[11] i. These authors also emphasize that using least squares cross-validation to determine the bandwidths allows to automatically discriminate between relevant

[9]See Stone (1985) or the monographs by Hastie and Tibshirani (1990) or Gu (2002) among others.

[10]Note that the stable and efficient GAM's estimation methodology developed in Wood (2004), which improves Wood (2000), is used to fit equation (31).

[11]Recall that given the use of 5-year data, the time dimension is of length four, which is small compared to the 25 countries observed each year.

and irrelevant regressors[12].

The relationship between the continuous predictors and the response in non or semiparametric regressions is usually reported graphically. Consequently, the results for specifications (31) and (32) are presented with partial regression plots. In that respect, we follow Maasoumi, Racine and Stengos (2007): if we wish to present the nonparametric regression of $GP_{i,t}$ on the continuous regressors $\mathbf{x}^c$ for EU15 countries, we plot

- $GP_{i,t}$ versus $E(GP_{i,t} \mid D_i = 1, \tilde{D}_t = \bar{t}, P_{i,t-T}, Y^{\star}_{i,t-T}, GY^{\star}_{i,t})$,
- $GP_{i,t}$ versus $E(GP_{i,t} \mid D_i = 1, \tilde{D}_t = \bar{t}, P^{\star}_{i,t-T}, Y_{i,t-T}, GY^{\star}_{i,t})$ and
- $GP_{i,t}$ versus $E(GP_{i,t} \mid D_i = 1, \tilde{D}_t = \bar{t}, P^{\star}_{i,t-T}, Y^{\star}_{i,t-T}, GY_{i,t})$,

where the upper star '$\star$' indicates that the variable is kept at its median level and $\bar{t}$ is a selected year. The same method is used for non-EU15 countries. Note that for the PLR model (31), the shapes for the nonparametric additive terms are similar for the pooled sample, for each country groupings and year up to an additive constant. For the fully nonparametric setting, interactions may drive to specific patterns depending on the levels of the discrete factors.

3.3 Distribution dynamics

While the regression framework allows a direct investigation of the theoretical reduced form equation, no information can be retrieved directly from the partial relationships regarding the evolution of the cross-country pollution gaps. As Barro and Sala-i Martin (2004) emphasize, a 'negative β' in equations of type (30) - often referred to as *conditional beta-convergence* - does not ensure a decrease in the cross-sectional dispersion of the level variable under investigation, even when the other conditioning factors are left out of the regression. Given the interest of the empirical literature for the dispersion dynamics of pollution across countries[13], we explore how the spatial distributions of the air pollutants' per capita emissions evolve over time with two standard approaches. The first one consists of comparing the spatial distribution of *past* emissions at regular time intervals without imposing constraints on the dynamic process. The researcher can check whether dispersion/disparities/entropy within the successive distributions reduces over time or whether the spatial density shapes become more peaked and concentrated around one or several modes (multi-polarization). The second approach suggested by Quah (1993, 1997) assumes that current (national) emission levels map into future ones according to a time-invariant and first order process of the form

[12] Irrelevant regressors are oversmoothed and the relationship becomes flat.

[13] See Ordás Criado and Grether (2009) for a recent review for carbon emissions or Bulte et al. (2007) for US state-level data for NOX and SOX emissions, among others.

$$\phi_{t+\tau}(z) = \int_0^\infty g_\tau(z|x)\phi_t(x)dx \tag{33}$$

where x and z denote the per capita emissions at times t and $t+\tau$, with $\tau > 0$, $\phi_t()$ is the spatial density function in t, $g_\tau(z|x)$ is the transitional (conditional density) operator and $\phi_{t+\tau}(z)$ is the resulting emissions' spatial distribution τ years later. In the latter framework, the additional structure imposed on the dynamic process allows to forecast spatial distributions based on the transitional law $g()$ and to compute long run (or steady state) distributions. Johnson (2000) proposes to convert the conditional density $g()$ into probabilities by discretizing the distribution support and to infinitely iterate the resulting conditional probability matrix to get the long run distributions in a business-as-usual scenario. In this paper, we use $\tau = 5$ and estimate the conditional density

$$\psi(y|\tilde{\mathbf{x}}^d, \tilde{\mathbf{x}}^c), \tag{34}$$

with the kernel estimator for mixed data proposed by Li and Racine (2003), where y is the dependent variable, ψ represents the conditional density, $\tilde{\mathbf{x}}^d$ and $\tilde{\mathbf{x}}^c$ are discrete and continuous predictors. More precisely, the comparative statics exercise consists in evaluating $\bar{\psi} = \psi\,(x \mid D_i,\, D_t)$, i.e. per capita emissions' densities conditional on the EU15 status (0/1) and years 1980, 1985, 1990, 1995, 2000 and 2005 (both treated as factors). The dynamics approach by Quah requires estimating the transitional operator given by $\tilde{\psi} = \psi\,(z \mid x,\, D_i)$ over the whole period. We further evaluate $\tilde{\psi}$ separately over the time horizons 1985-2005, 1990-2005, 1995-2005 and 2000-2005 to control for the time invariance hypothesis of $g()$. Therefore several long-run distributions, resulting from the latter time periods, are reported along with the spatial kernel fit for the terminal year 2005 for comparison purposes. Note that we employ in all cases the simplest version of the mixed estimator (normal reference rule-of-thumb[14] for the continuous explanatory variables and unsmoothed discrete factors), which corresponds to the frequency approach. This avoids undersmoothed patterns in the presence of up to 25% of outliers in the sample without modifying the fundamental distributional trends.

4 Data

A consistent empirical testing of the optimal pollution-GDP relationship (29) imposes two basic requirements regarding the pollution data: (i) their negative impact on collective welfare needs to be linked to their flow (and not to their stock) and (ii) (transboundary) regulatory mechanisms must be at work to enforce (potentially optimal) defensive measures. Since the eighties, the European states have been particularly pro-active in fighting atmospheric pollution, and more particularly two

[14]See Silverman (1986, p.47).

acidifying gases' emissions: nitrogen oxides (NOX) and sulphur oxides (SOX). The Helsinki 1985 and the posterior Olso 1994 and Goteborg 1999 Protocols bound about twenty European countries to reduce substantially their sulfur emissions through a variety of mechanisms. NOX emissions experienced similar early control initiatives across Europe through the 1988 Sofia Protocol or the Large Combustion Plant European Directive (2001/80/EC)[15]. Therefore exploring the presence of the defensive and scale effects for per capita NOX and SOX emissions with a European panel of countries covering the post 1980 period appear as a natural step to test the pollution convergence equation.

Our database is a balanced panel of 25 European (and Asian) countries that covers the 1980-2005 period. We use GDP and population series from Maddison (2008) while the NOX and SOX emissions come from the EMEP-CEIP database WebDab and correspond to series used in the EMEP models. These pollution data are based on officially reported emissions, but inconsistent/missing observations are corrected and/or gap-filled[16]. The European Monitoring and Evaluation Programme (EMEP) is a protocol signed in 1984 under the Convention on Long-Range Transboundary Air Pollution (CLRTAP) which requires that parties report on several air pollutant emissions to the treaty secretariat. Since January 2008, the EMEP Centre on Emission Inventories and Projections (CEIP) operates the EMEP emission database (WebDab), which records anthropogenic and natural emissions for a large variety of air pollutants (acidifying/eutrophying compounds, ozone precursor, heavy metals and particulate matters). Our analysis focuses on NOX and SOX emissions derived from human activities and ignores those occurring in natural environments without human influence. Sulfur and nitrogen oxides are well-known to have a large negative impact on human health and natural ecosystems[17]. Through the reaction with other substances, NOX and SOX emissions cause lung diseases; they modify land and water ecosystems and generate acid rains that affect nature as well as buildings, cars or historical monuments. NOX particles are essentially emitted by road transportation, other mobile sources, and electricity generation, which correspond to the so-called Selected Nomenclature for Air Pollution (SNAP97) categories S7, S8 and S3 in the CORINAIR-EMEP classification. SOX emissions are mainly linked to combustion processes at the industry and plants' level (SNAP97 sectors S1 to S3).

Figure 2 displays the national series on per capita SOX and NOX emissions as well as GDP per capita. Solid lines designate the historical EU members EU15, while the most recent Eastern EU members and the non-EU members are pre-

[15]The reader interested in the effectiveness of these Protocols in mitigating pollution may refer to Murdoch, Sandler and Sargent (1997) (Sofia and Helsinki Protocols), Finus and Tjotta (2003) (Oslo's) or Bratberg, Tjotta and Oines (2005) (Sofia's).

[16]More details are provided in the CEIP technical report Inventory Review 2008, see http://www.ceip.at/review-process/review-2008/.

[17]See the US Environmental Protection Agency at http://www.epa.gov/air/urbanair/.

sented with dash lines and are called the non-EU15 group[18]. The left graph shows that all per capita GDP series are upward trended and that all non-EU15 countries but Switzerland have significantly lower per capita GDP levels. There is no clear evidence of a decreasing gap in per capita GDP either within or between these two groups over the 1980-2005 period. By contrast, many of the downward sloping per capita NOX and SOX series seem to stabilize at some point and/or to converge across the whole sample. These rough patterns suggest the presence of two possible distinct groupings of countries, which may display different behaviors in the econometric analysis.

5 Results

Tables 1 and 2 contains the regression results for NOX and SOX respectively. Column (A) tests the 'short β-convergence' OLS specification for per capita pollution often employed in the empirical literature[19]. Columns (B) and (C) are OLS parametric fits of our theoretical equation (29), which also represents a 'β-type' convergence regression for pollution, conditional upon the levels and growth rates of per capita GDP. Column C is an Instrumental Variables (IV henceforth) estimation which controls for potential endogeneity bias. Columns (D) shows the linear part of the pooled IV semiparametric while column (E) only displays the R-squared of the nonparametric estimates. Graphical devices (Figures 3 and 6 for NOX and SOX respectively) complete the former results by displaying the linear, partially linear and fully nonparametric partial relationships, for year $\bar{t} = 1985$ and by EU15 status keeping all other continuous factors at their median levels.

Starting with the results for the NOX emissions in Table 1, we notice that all parametric regressions display heteroscedastic errors as the null of homoscedasticity is overwhelmingly rejected with the LM-test. The coefficients' standard deviation for the parametric fits are White-corrected while those of the linear part of the semiparametric model rely on a bayesian approach[20]. We can see that all coefficients are significant, at the 1% level for the vast majority, across all models. Moreover, taking into account the GDP variables (Y and GY) improves the explanatory power of the models compared to the parsimonious regression (A) as the adjusted R-squared increases substantially for all alternative models. Model (A) establishes 'β convergence' for NOX emissions with significantly larger growth rates in the EU15 countries. The conditional specifications (B) and (C) yield similar results with respect to past pollution levels when the role of GDP enters into

[18]The EU15 and non-EU15'*' countries are: Albania*, Austria, Belgium, Bulgaria*, Czechoslovia*, Denmark, Finland, France, Former Yugoslavia*(without Serbia and Montenegro), Former USSR*, Germany, Greece, Hungary*, Ireland, Italy, Netherlands, Norway, Poland*, Portugal, Romania*, Spain, Sweden, Switzerland*, Turkey*, United Kingdom.

[19]See List (1999), Strazicich and List (2003), Nguyen Van (2005) or Brock and Taylor (2004b).

[20]See Wood (2006, Ch.4.8 and 4.9) for further details.

play. Our estimates display a significant positive scale effect linked to GDP growth as the theoretical model predicts. The effect of past GDP is negative in models (B) as well as in the IV counterpart (C). When the specification test of Hsiao et al. (2007) is applied to the OLS fits at the bottom of columns, all the parametric models are rejected at the 1% level, which means that they are misspecified. Consequently, the flexible approaches (D) and (E) are expected to depict nonlinearities as well as potentially different patterns, specific to EU15 membership for the fully nonparametric model. The PLR estimates in column (D) confirm that time dummies matter and that the greater flexibility introduced for the continuous regressors clearly increases the models' explanatory power. From that perspective, the fully nonparametric model in column (E) appear as being even superior as it captures 88% of the total variance. Therefore, we put the emphasis on the fully nonparametric fits and report the partial PLR and OLS estimates for the continuous regressors on the same partial regression plots for comparison purposes.

Firstly, we observe in the upper plots of Figure 3 that the defensive effect linked to past pollution is confirmed for the EU15 and for the non-EU15 countries with the nonparametric partial fits. The least-square cross-validation methodology employed to determine the bandwidths does not detect significant departure from linearity for that partial relationship. Secondly, the scale effect linked to GDP growth is positive as expected, with larger partial elasticities for GDP growth in the EU15 economies. The effect of past GDP levels on the posterior growth rates of per capita NOX emissions is more ambiguous as the confidence interval includes the zero over large portions of the support for both groups. However, we clearly see that the latter variable does capture nonlinear effects that may have been neglected within specification (28). Finally, note that, the confidence interval surrounding the non-EU15 fits are larger. In sum, the path followed since 1985 by the NOX per capita emissions seems compatible with the predicted optimal pattern associated with pollution convergence, but with a stronger evidence holding within the EU15 countries[21].

We now turn to the distributional analysis for the NOX per capita emissions in Figure 4. The left-hand-side plots indicate that the rather spread spatial distribution in 1980 tends to become single-peaked, more symmetric and concentrated over a lower and tighter emissions' support as time goes by. This holds for the EU15, non-EU15 and for the pooled sample. The long run distributions (right-hand-side plots) are also predicted to converge toward a peaked and unimodal shape when we use a transitional law covering the whole 1980-2005 time horizon. That tendency is reinforced over time for the EU15 countries and the pooled sample at a lesser degree as the ergodic distributions computed with transitional laws estimated over nearer-term time horizons tend to be slightly left-shifted, sharper and more concen-

[21]The partial relationships for the fully nonparametric regressions being potentially different for each level of the time factor $\bar{t} = \{1985, 1990, 1995, 2000\}$, note that the defensive effect tends to become flatter for NOX emissions as time goes by. The partial regression fits at each time level are available upon request.

trated. We conclude that the converging pollution relationship is also associated with cross-country convergence in NOX per capita emissions.

Additional exploratory analysis can be carried out with equation (29) by reversing the correlation structure between the dependent pollution growth variable ($GP_{i,t}$) and the GDP growth counterpart ($GY_{i,t}$). In this case, we get a function that explores how GDP growth rates are correlated with initial GDP levels, initial pollution levels and pollution growth. The interest of that formulation is to check empirically to what extent the estimated pollution growth equation is compatible with 'β-convergence' in per capita GDP conditional on environmental quality. Indeed, the exact same analysis conducted for pollution growth can be applied to GDP growth[22]. The OLS estimates for the latter parametric model being rejected at the 5% cutoff with the specification test[23], we present directly the fully nonparametric regressions with their usual (semi)parametric benchmark on Figure 5. We can see on the upper panel that GDP growth rates and Initial GDP levels display a partial positive correlation in the EU15 and the non-EU15 countries, which points toward rejecting the conditional β-convergence relationship in per capita GDP in Europe over that period. Figure 9 further shows that cross-country GDP per capita differences within the EU15 and the non-EU15 countries are rather persistent over the years 1980-2005. In both cases the yearly spatial densities shift to the right and become slightly less peaked, while the long run distributions appear as being more spread than in 2005, whatever the time span considered. For the pooled sample, the yearly cross-section densities tend toward bi-modality but the Markov approach rather forecasts unimodal and left-skewed ergodic shapes with a large basis in the long run. Back to the middle and bottom panels of Figure 5, the nonparametric fits indicate that past per capita emission levels neither penalize nor favor economic growth in the EU15 countries but they are positively linked to GDP growth in the non-EU15 countries. This may indicate that controlling for pollution may hurt growth in these countries. Regarding the partial GDP growth - pollution growth link, the positive correlation found with specification (29) also holds here but the relationship is rather of linear type.

Finally, the SOX results depart from the NOX ones in several aspects. First, we can see in Table 2 that the explanatory power of the models is lower and that no heteroskedasticity is detected in the linear models. Second, while specifications (B) and (C) are rejected, the most parsimonious linear model in column (A) is considered correctly specified, even if it captures a very low proportion of the total variance. Third, we do find with the nonparametric fits in Figure 6 a defensive effect linked to initial pollution levels for SOX, but the effect of initial GDP is clearly negative and linear. Fourth, while a positive scale effect in GDP growth is found for the EU15 countries for SOX, a U-inverted shape characterizes the fully nonparametric partial relationship for the non-EU15[24]. Regarding the distri-

[22] In that setting, the IV estimates use $P_{i,t-1}$ as instrument for $P_{i,t}$ and $GP_{i,(t-1)-T}$ for $GP_{i,t-T}$.

[23] Detailed results for these complementary regressions are available upon request.

[24] The partial fully nonparametric fits are very similar for alternative levels of the time factor

bution dynamics for SOX, Figure 7 shows that there are clear tendencies toward cross-country convergence, with yearly spatial densities becoming single-peaked and more concentrated over a lower and tighter emissions range and projected long run densities for the EU15, nonEU15 and pooled samples which tend toward cross-sectional absolute convergence. Finally, we observe in Figure 8 that we find 'β-divergence' in GDP conditional on the SOX variables, no partial relationship between past pollution and subsequent GDP growth, a positive partial link between GDP growth and pollution growth. Similarly to the NOX case, these results for SOX emissions confirm that 'β-convergence' in pollution, conditional on GDP, can perfectly co-exist with 'β-divergence' in income, conditional on pollution, and that increasing or stable gaps in per capita GDP between countries, in a context of rising per capita GDP levels, are fully compatibles with decreasing pollution gaps, in a context of decreasing pollution levels.

6 Conclusion

This paper investigates the income-pollution link within a β-type regression setting. A dynamic relationship involving pollution growth, past pollution levels, past GDP levels and GDP growth is derived from a growth model *à la* Ramsey with optimal pollution control through investment in clean technologies. This test-equation predicts a positive effect of GDP growth on pollution growth and an offsetting effect, linked to initial pollution levels, that captures defensive expenditures and emissions-reducing technological progress along the steady state. Estimates based on parametric, semiparametric as well as fully nonparametric regressions are computed for a panel of 25 European countries on per capita NOX and SOX emissions spanning the years 1980 to 2005. We find evidence of the two predicted effects influencing the pollution dynamic: a clear scale effect linked to GDP growth and a negative effect captured through the impact of the past pollution level component. It also appears that β-convergence in pollution, conditional on the levels and dynamics of per capita GDP, is empirically fully compatible with β-divergence in per capita GDP levels, conditional on the level and dynamics of environmental quality. We further show with a distribution analysis that the gaps in NOX and SOX per capita emissions between countries reduce over time while those in per capita income raise or remain steady. Therefore, reducing income disparities does not appear as being a prerequisite to decreasing pollution gaps between the European countries. Achieving closer per capita pollution levels can indeed take place at the same time or even before income equalization occurs.

($\bar{t} = \{1990, 1995, 2000\}$) and are available upon request.

Tables & Figures

Table 1: Regression results : NOX pollution growth *vs.* initial pollution levels and GDP.

	Parametric models			Non/semipa. models	
	Ordinary LS			PLR fit$^{(a)}$	NP fit$^{(b)}$
Variables	(A)	(B)	(C)	(D)	(E)
constant	0.137***	0.127***	0.126***	-0.027***	-
d1990	-0.048***	-0.026***	-0.024***	-0.021***	-
d1995	-0.034***	-0.038***	-0.038***	-0.034***	-
d2000	-0.035***	-0.037***	-0.035***	-0.034***	-
EU15	0.032***	0.019*	0.030***	0.059***	-
$P_{i,t-T}$ (β)	-0.043***	-0.037***	-0.033***	-	-
$Y_{i,t-T}$ (λ)		-0.010*			
$Y_{i,(t-1)-T}$ (λ)			-0.014**	-	-
$GY_{i,t-T}$ (θ)		0.653***			
$GY_{i,(t-1)-T}$ (θ)			0.590***	-	-
N	100	100	100	100	100
R2/R2 adj.	0.39/0.36	0.56/0.53	0.56/0.53	0.71/0.66	0.88/-
F-stat	12.3***	16.7***	16.9***		
Heterosced.$^{(c)}$	16.9***	17.6***	18.9***		
P(Specific.)$^{(d)}$	0.000	0.000	0.002		

Notes: ***, ** and * denote the 1%, 5% and 10% significance levels. All computations made in *R.2.9.1.* (a): The PLR estimation (D) is computed with the *gam* function from the *mgcv, v.1.5-5* package with the default options. (b): The nonparametric mixed fit (E) is estimated with the *npbwreg* function from the *np, v.0.30-3* package with options *bwmethod="ls.cv", regtype="ll", ukertype="liracine", nmulti=50.* Detailed bandwidths are available upon request. (c) : 'Heterosced.' is the heteroscedasticity LM-test by Breusch and Pagan (1979), computed with the variance estimator proposed by Koenker (1981), robust to departure from normality. The latter statistic is χ^2-distributed, with d.f. = nb. of regressors (constant excluded). (d): 'P(Specific.)' stands for the probability associated to the nonparametric specification test by Hsiao et al. (2007) for continuous and discrete data models (see the function *npcmstest*, package *np, v.0.30-3*. The latter probability is based on 399 bootstrap's replications.

Table 2: Regressions results : SOX pollution growth *vs.* initial pollution levels and GDP.

	Parametric models			Non/semipa. models	
	Ordinary LS			PLR fit[(a)]	NP fit[(b)]
Variables	(A)	(B)	(C)	(D)	(E)
constant	0.014	0.152***	0.143***	-0.079***	-
d1990	-0.048**	-0.024	-0.022	-0.020	-
d1995	-0.034	-0.041**	-0.041**	-0.042**	-
d2000	-0.019	-0.020	-0.020	-0.028	-
EU15	-0.018	0.030	0.029	0.066**	-
$P_{i,t-T}$ (β)	-0.012	-0.021***	-0.019**	-	-
$Y_{i,t-T}$ (λ)		-0.062***			
$Y_{i,(t-1)-T}$ (λ)			-0.062***	-	-
$GY_{i,t-T}$ (θ)		0.660**			
$GY_{i,(t-1)-T}$ (θ)			0.568**	-	-
N	100	100	100	100	100
R2 / R2 adj.	0.09/0.04	0.28/0.22	0.27/0.22	0.48/0.40	0.60/-
F-stat	1.88	5.02***	4.94***		
Heterosced.[(c)]	5.64	8.55	8.00		
P(Specific.)[(d)]	0.303	0.002	0.005		

Notes: ***, ** and * denote the 1%, 5% and 10% significance levels. All computations made in *R.2.9.1.* (a): The PLR estimation (D) is computed with the *gam* function from the *mgcv, v.1.5-5* package with the default options. (b): The nonparametric mixed fit (E) is estimated with the *npbwreg* function from the *np, v.0.30-3* package with options *bwmethod="ls.cv", regtype="ll", ukertype="liracine", nmulti=50.* Detailed bandwidths are available upon request. (c) : 'Heterosced.' is the heteroscedasticity LM-test by Breusch and Pagan (1979), computed with the variance estimator proposed by Koenker (1981), robust to departure from normality. The latter statistic is χ^2-distributed, with d.f. = nb. of regressors (constant excluded). (d): 'P(Specific.)' stands for the probability associated to the nonparametric specification test by Hsiao et al. (2007) for continuous and discrete data models (see the function *npcmstest*, package *np, v.0.30-3*. The latter probability is based on 399 bootstrap's replications.

Figure 1: Phase diagram of capital vs pollution.

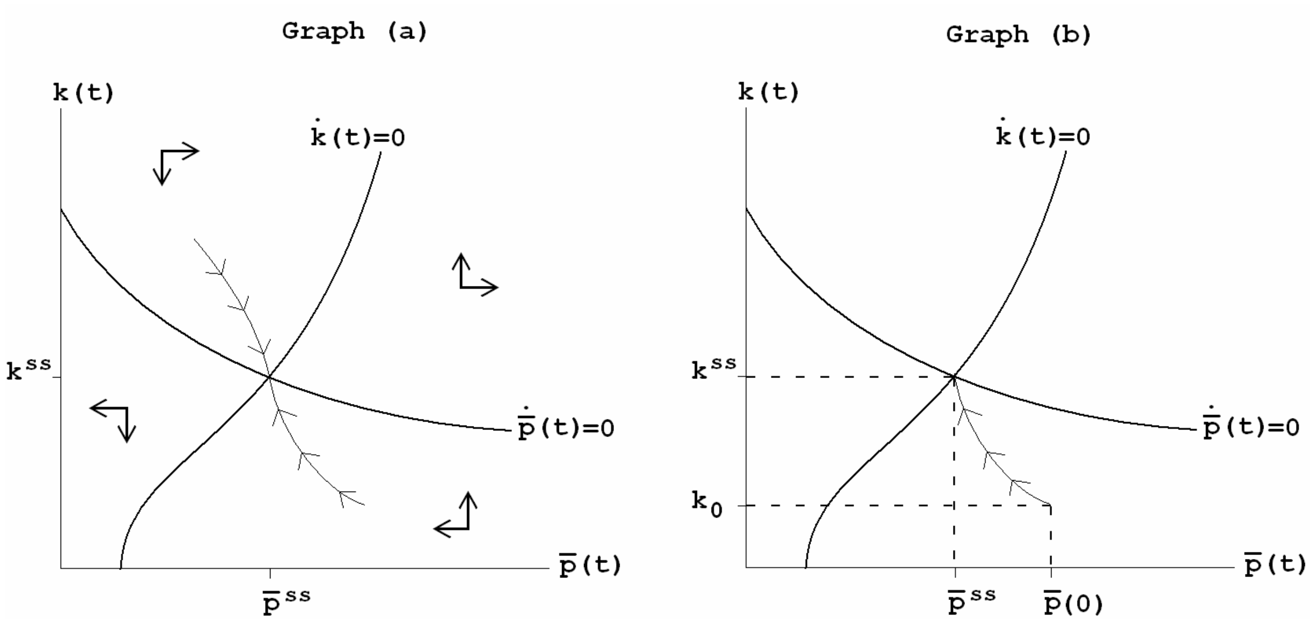
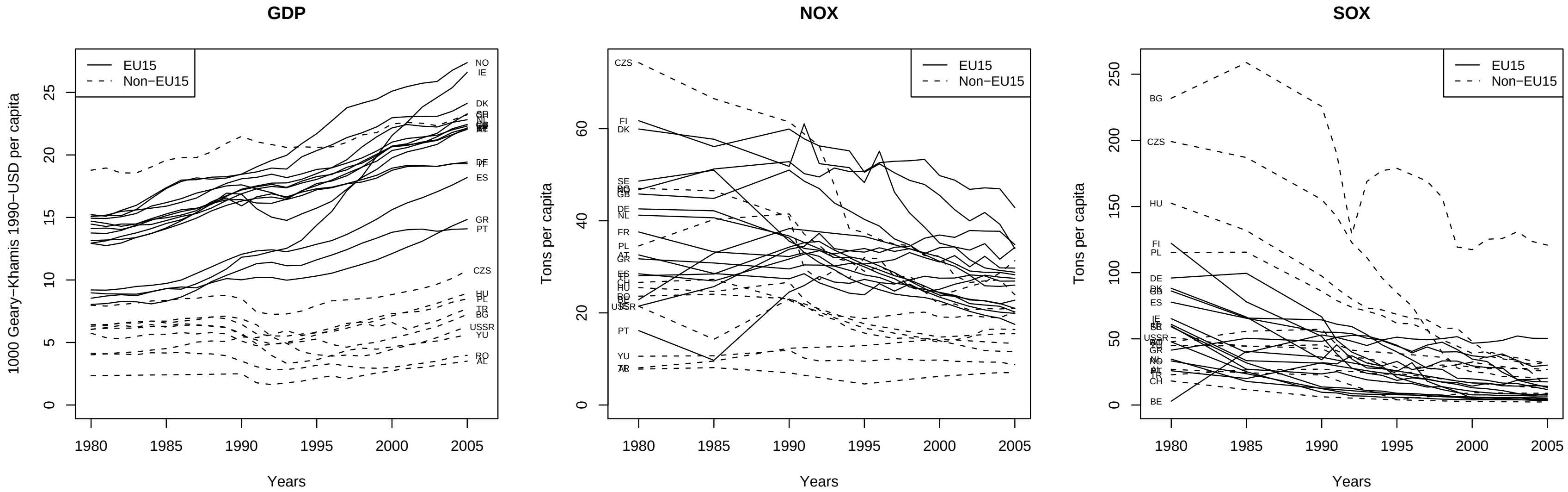
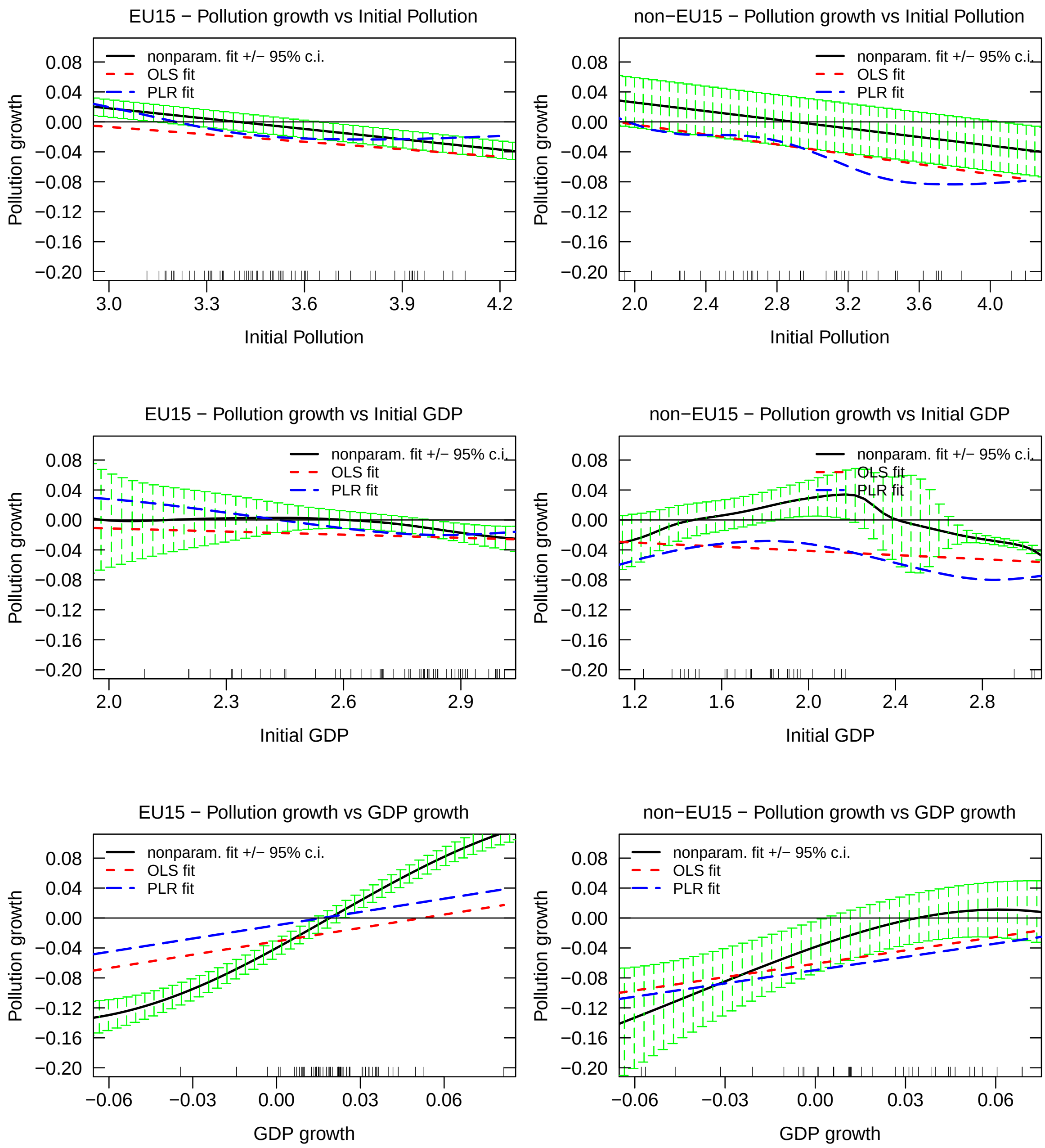


Figure 2: Per capita levels of GDP, SOX and NOX. European countries, national trends 1980-2005.



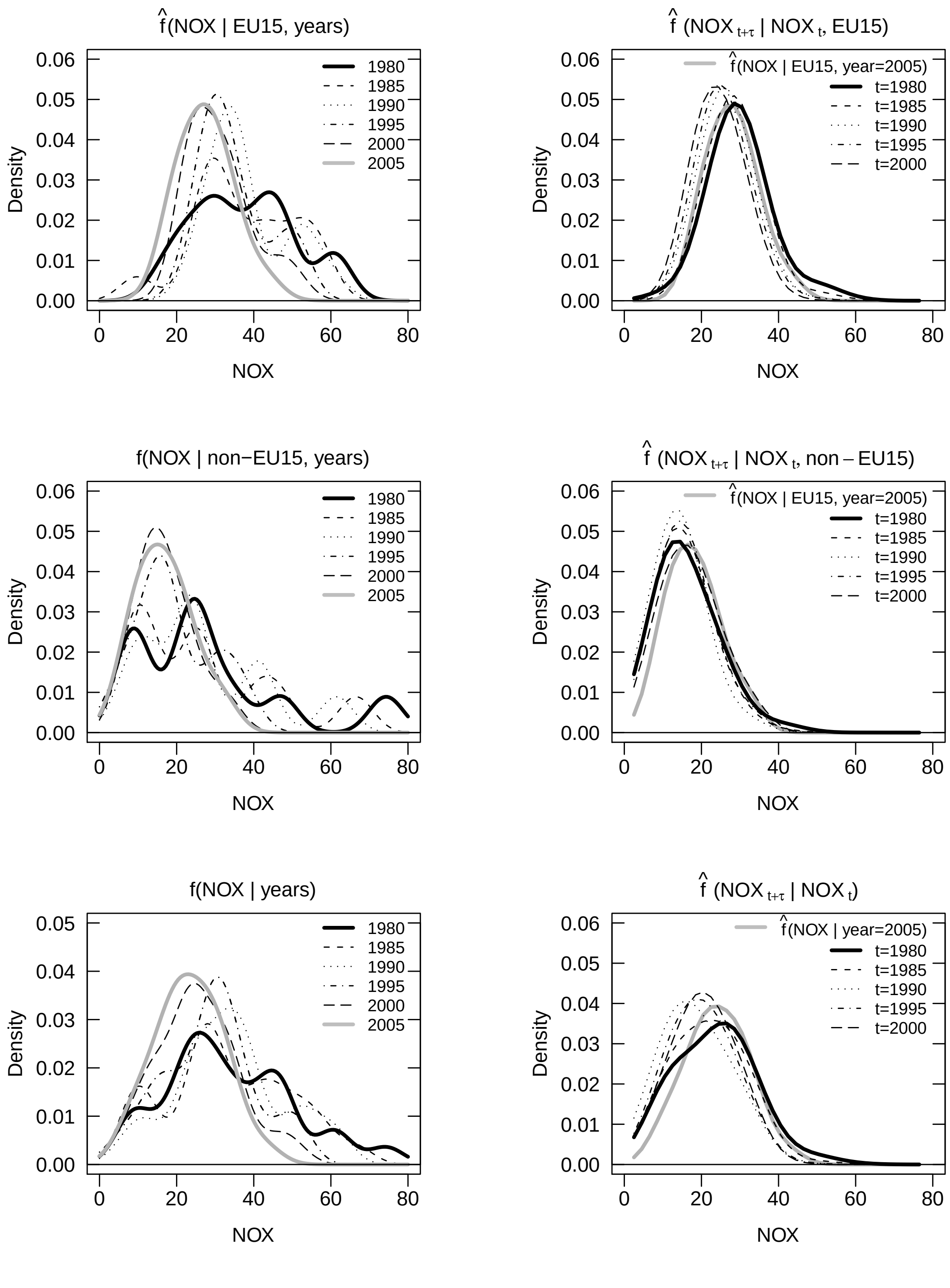
Note: GDP and population figures come from Maddison (2008). Emissions come from the Centre on Emission Inventories and Projections and correspond to the series used in EMEP models. Emissions for the 1980-1990 period were available only for 1980 and 1985.

Figure 3: Nonparametric partial regressions by EU15 status : NOX pollution growth *vs.* initial pollution levels and GDP.



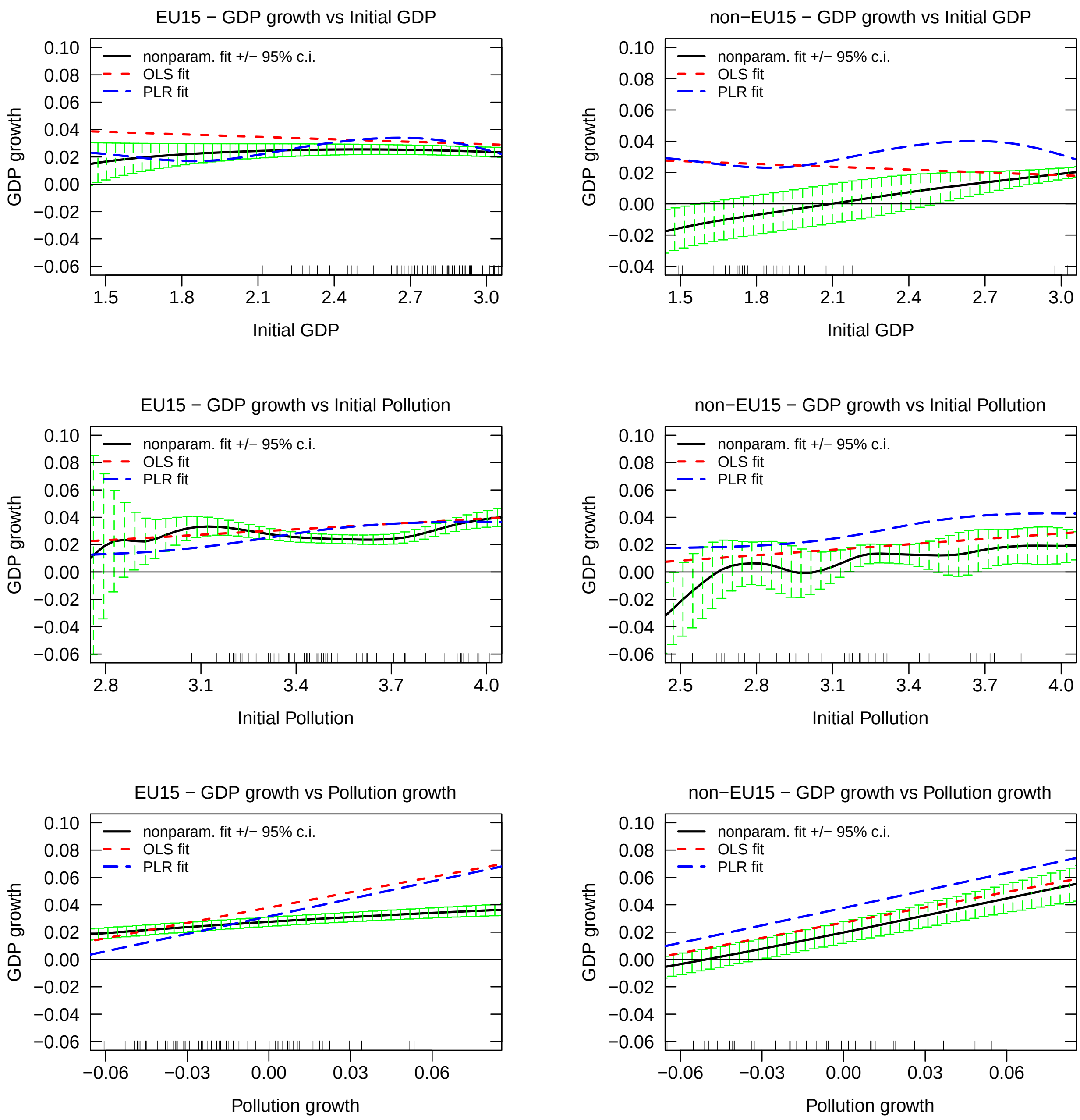
Notes: Nonparametric regressions based on Racine and Li (2004). Partial regression estimates computed with the function *npregbw* from library *np, v.0.30-0* with *bwmethod="ls.cv", regtype="ll", ukertype="liracine", nmulti=50.* Confidence intervals are based on asymptotic standard errors.

Figure 4: Distribution dynamics for NOX : yearly spatial densities vs long run distributions by EU15 status.



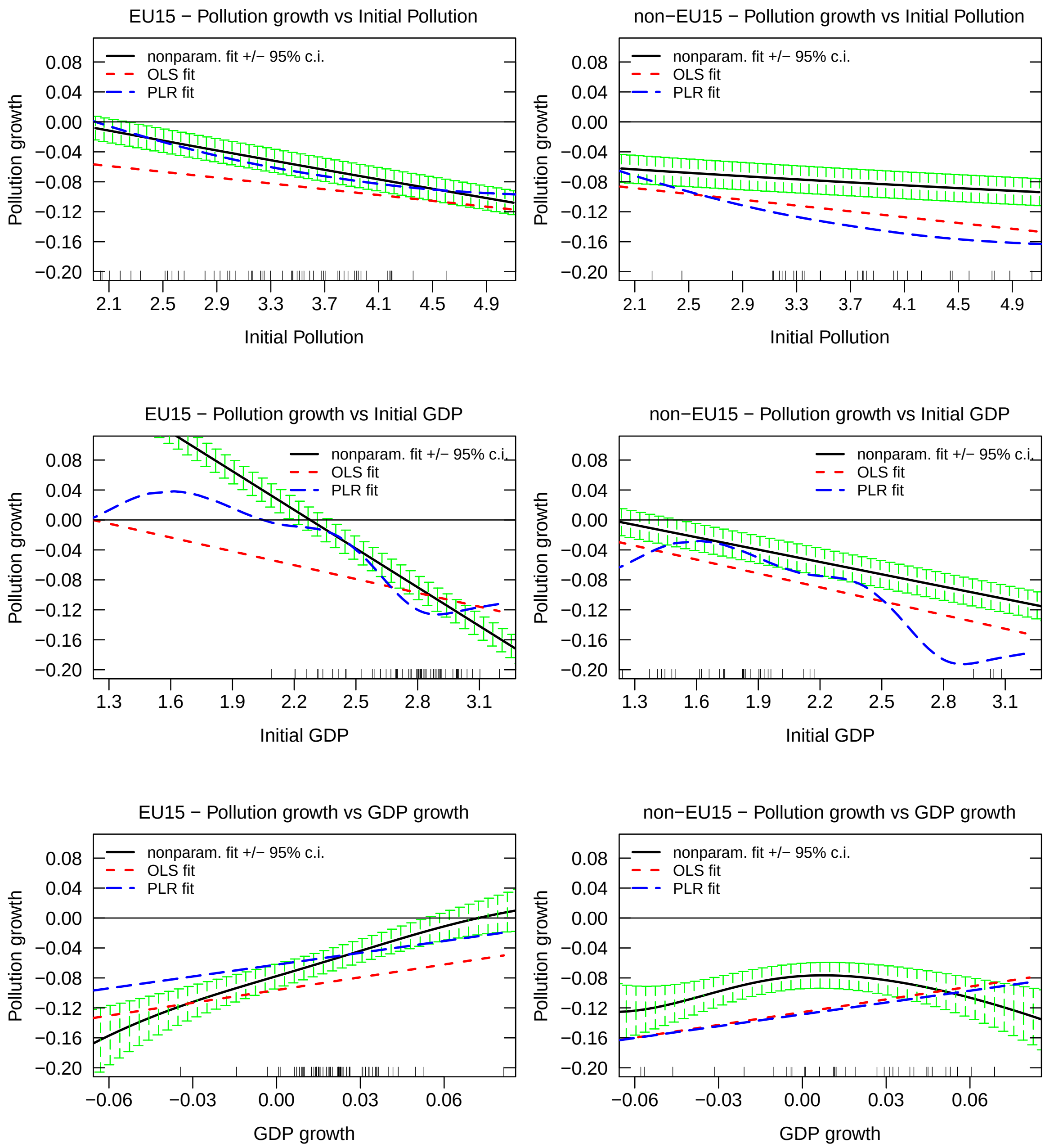
Notes: Conditional densities computed with the function *npcdensbw* from the library *np, v.0.30-3* with *bwmethod="normal-reference", bwtype="fixed",ukertype="liracine"*. EU15 and years are treated as factors.

Figure 5: Nonparametric partial regressions by EU15 status : GDP growth *vs.* initial GDP levels and NOX emissions.



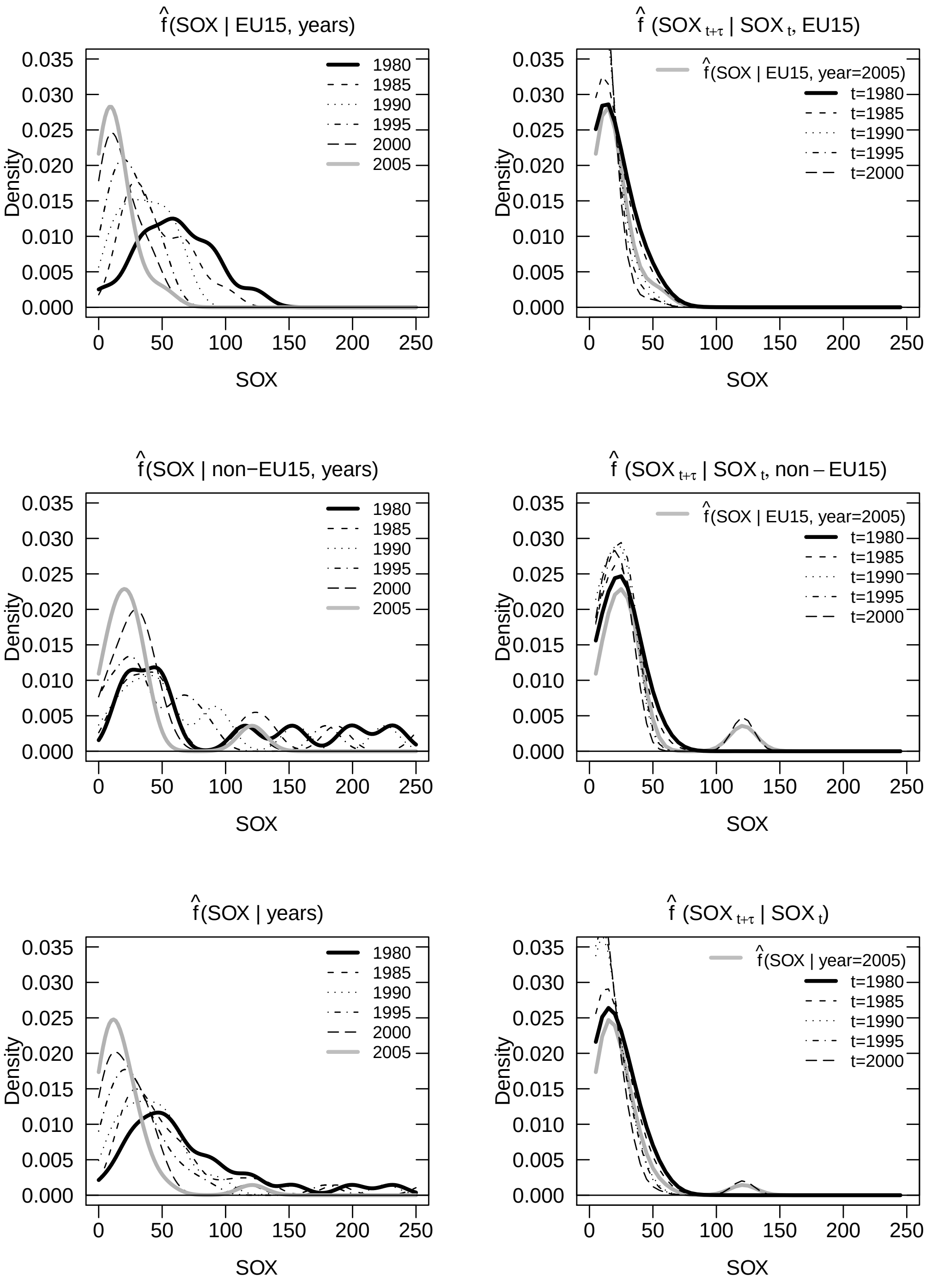
Notes: Nonparametric regressions based on Racine and Li (2004). Partial regression estimates computed with the function *npregbw* from library *np, v.0-30.0* with *bwmethod="ls.cv", regtype="ll", ukertype="liracine", nmulti=50.* Confidence intervals are based on asymptotic standard errors.

Figure 6: Nonparametric partial regressions by EU15 status : SOX pollution growth *vs.* initial pollution levels and GDP.



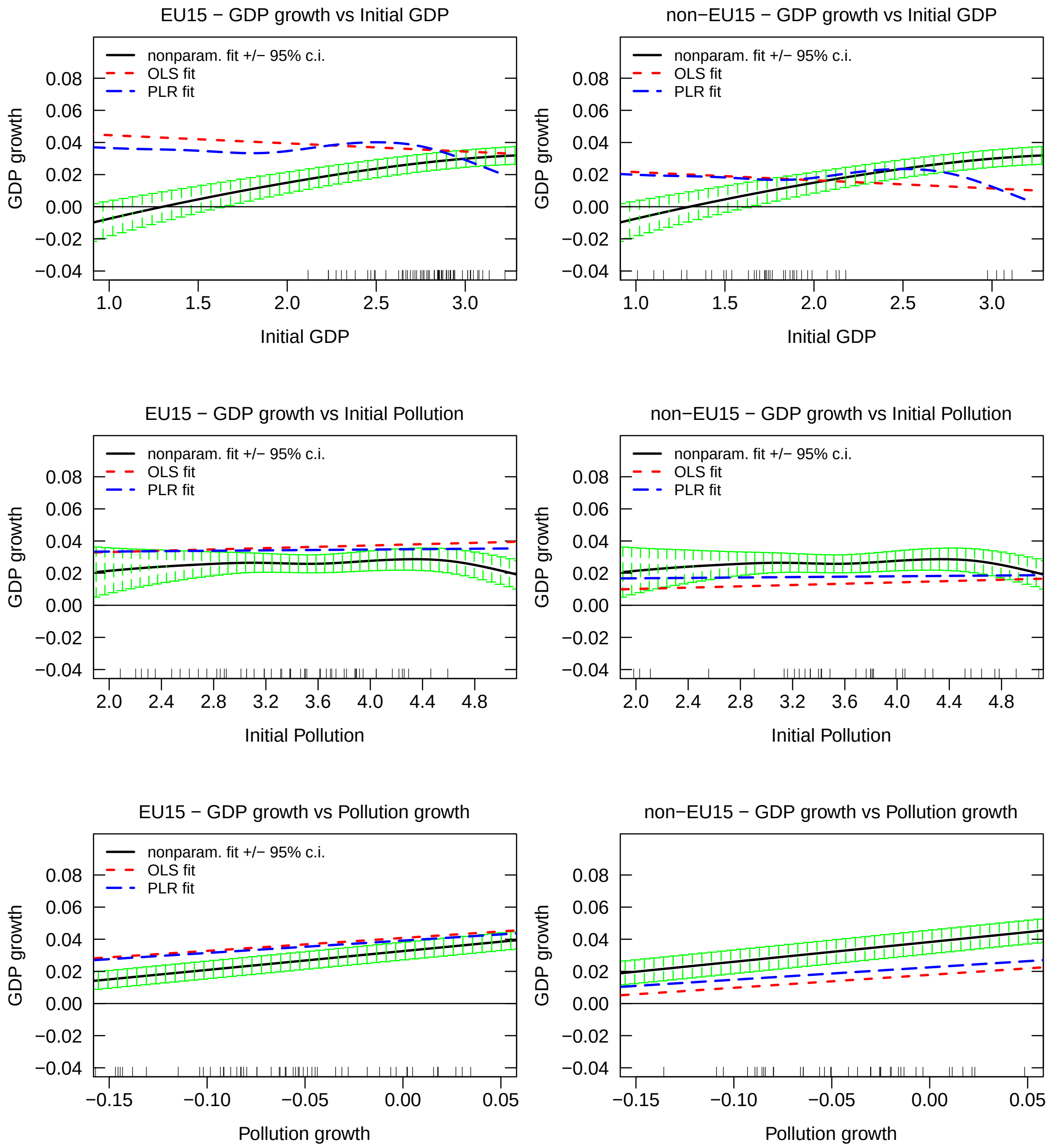
Notes: Nonparametric regressions based on Racine and Li (2004). Partial regression estimates computed with the function *npregbw* from library *np, v.0.30-0* with *bwmethod="ls.cv", regtype="ll", ukertype="liracine", nmulti=50.* Confidence intervals are based on asymptotic standard errors.

Figure 7: Distribution dynamics for SOX : yearly spatial densities vs long run distributions by EU15 status.



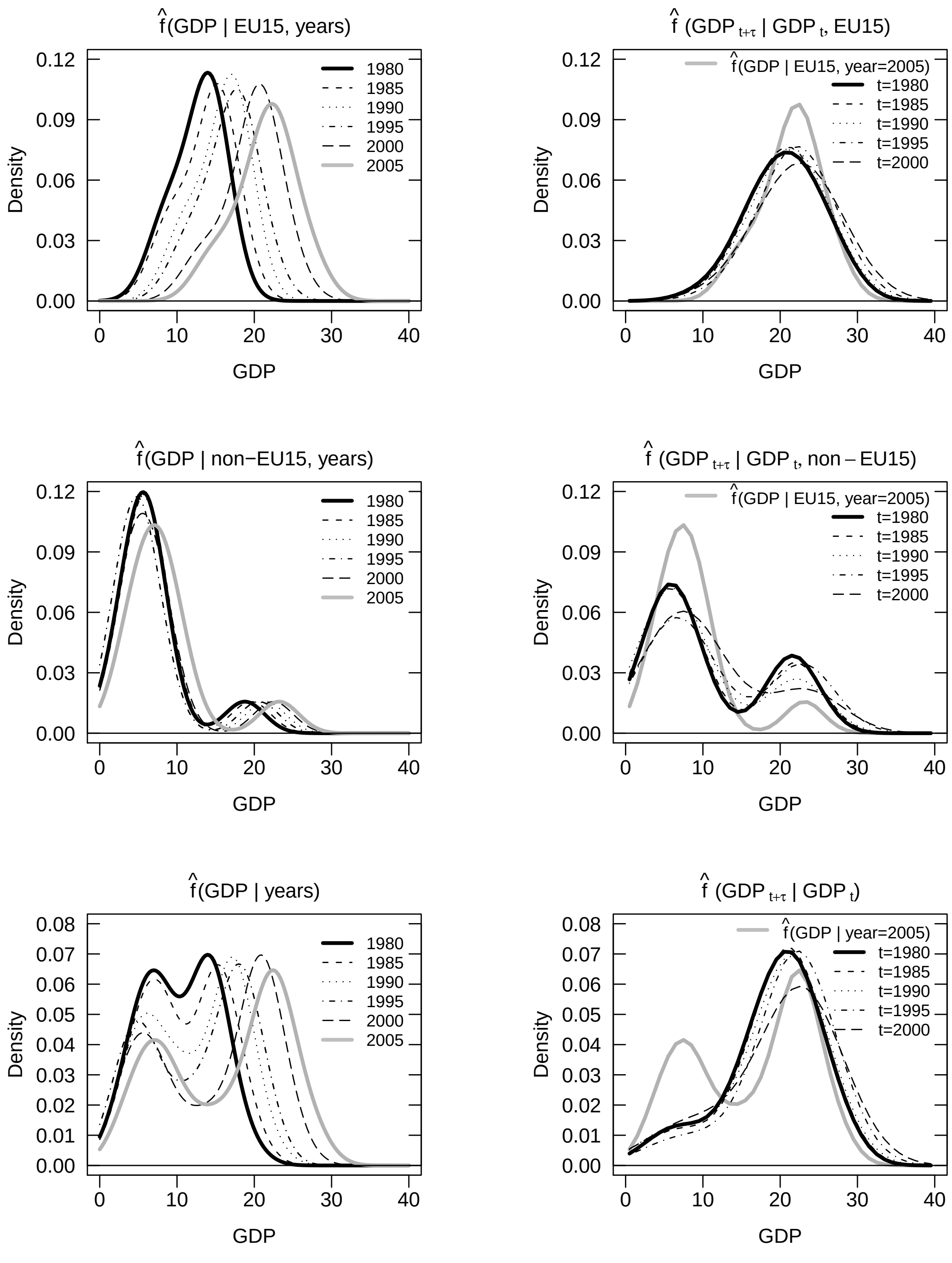
Notes: Conditional densities computed with the function *npcdensbw* from the library *np, v.0.30-3* with *bwmethod="normal-reference", bwtype="fixed",ukertype="liracine"*. EU15 and years are treated as factors.

Figure 8: Nonparametric partial regressions by EU15 status : GDP growth *vs.* initial GDP levels and SOX emissions.



Notes: Nonparametric regressions based on Racine and Li (2004). Partial regression estimates computed with the function *npregbw* from library *np, v.0.30-0* with *bwmethod="ls.cv", regtype="ll", ukertype="liracine", nmulti=50*. Confidence intervals are based on asymptotic standard errors.

Figure 9: Distribution dynamics for GDP : yearly spatial densities vs long run distributions by EU15 status.



Notes: Conditional densities computed with the function *npcdensbw* from the library *np, v.0.30-3* with *bwmethod="normal-reference", bwtype="fixed",ukertype="liracine"*. EU15 and years are treated as factors.

Appendix

Derivation of (10)-(11). Substituting $\bar{c} = cB$ and $\bar{p} = pB$ in the utility function (6), the optimal control problem consists of finding the optimal path

$$\{c(t), \chi(t)\}_{t=0}^{\infty} = \arg\max \int_0^{\infty} U(c(t)B(t), p(t)B(t))\, e^{-\rho t} dt$$

subject to (8)-(9). The current-value Hamiltonian associated with this problem is (omitting time-arguments for simplicity)

$$\begin{aligned} H(c,p,k) &= \sigma \ln(cB) - \varsigma (pB)^{\theta} + \lambda^k [f(k)(1-\chi) - c - (\delta+n+\pi)k] + \\ &\quad + \lambda^p [\Omega (1-\chi)^{\varepsilon} f(k) - p], \end{aligned}$$

where λ^k is the dynamic multiplier associated with the dynamic constraint (8), and λ^p is the Lagrange multiplier associated with the current-time constraint (9). Assuming an interior solution, the necessary conditions for optimality are

$$\begin{aligned} H_c &= 0 \qquad \lambda^k = \sigma c^{-1}, && (35)\\ H_p &= 0 \qquad \lambda^p = -\varsigma\theta p^{\theta-1}B^{\theta}, && (36)\\ H_{\chi} &= 0 \qquad \lambda^k = -\lambda^p \Omega\varepsilon (1-\chi)^{\varepsilon-1} = 0, && (37) \end{aligned}$$

together with the co-state equation $H_k = \rho\lambda^k - \dot{\lambda}^k$ and the transversality condition on capital, that read

$$\begin{aligned} \dot{\lambda}^k/\lambda^k &= \rho+\delta+n+\pi - f_k(1-\chi) - \left(\lambda^p/\lambda^k\right)\Omega(1-\chi)^{\varepsilon} f_k, && (38)\\ 0 &= \lim_{t\to\infty} \lambda^k(t)\, k(t)\, e^{-\rho t}, && (39) \end{aligned}$$

respectively. From (37), we can substitute $\lambda^k/\lambda^p = -\Omega\varepsilon(1-\chi)^{\varepsilon-1}$ in (38) to obtain

$$\dot{\lambda}^k/\lambda^k = \rho+\delta+n+\pi - f_k(1-\chi)\left(1-\varepsilon^{-1}\right). \qquad (40)$$

Time-differentiation of (35) yields $\dot{c}/c = -\dot{\lambda}^k/\lambda^k$, which can be plugged into (40) to obtain (10). Combining (35), (36) and (37) to eliminate λ^k and λ^p, we obtain

$$\Omega\varepsilon(1-\chi)^{\varepsilon-1}\varsigma\theta p^{\theta-1}B^{\theta} = \sigma c^{-1},$$

where we can substitute $p = \Omega(1-\chi)^{\varepsilon} f(k)$ from (9) to obtain

$$(1-\chi)^{\varepsilon\theta-1} = \left[\frac{\sigma}{\varepsilon\varsigma\theta(\Omega B)^{\theta}}\right]\frac{f(k)^{1-\theta}}{c}. \qquad (41)$$

Since the product $\Omega(t)B(t) = \Omega_0 B_0$ is constant, the term in square brackets in (41) is constant over time. Hence, solving (41) for $(1-\chi)$ and defining $\Gamma \equiv \left[\varepsilon\varsigma\theta(\Omega_0 B_0)^{\theta}\sigma^{-1}\right]^{-1/(\varepsilon\theta-1)}$, we obtain (11). $\square$

Existence and uniqueness of the steady-state equilibrium. Imposing $\dot{c}=0$ in (14) and setting $k=k^{ss}$ and $c=c^{ss}$ in the resulting expression, we obtain equation (15). Substituting $\Phi\left(k^{ss},c^{ss}\right)$ by means of (12), and solving for c^{ss}, we obtain

$$c^{ss}=\left[f_k^{ss}\left(\Gamma/\tilde{\rho}\right)\left(1-\varepsilon^{-1}\right)\right]^{\varepsilon\theta-1}f\left(k^{ss}\right)^{-(\theta-1)}\equiv\psi_a\left(k^{ss}\right), \tag{42}$$

where we have defined the function $\psi_a\left(k^{ss}\right)$. Similarly, impose $\dot{k}=0$ in (8), set $c=c^{ss}$, and substitute $1-\chi^{ss}=\tilde{\rho}\left[f_k^{ss}\left(1-\varepsilon^{-1}\right)\right]^{-1}$ from (15) to obtain

$$c^{ss}=\frac{\rho f\left(k^{ss}\right)+\left(\delta+n+\pi\right)\left[f\left(k^{ss}\right)-f_k^{ss}k^{ss}\right]}{f_k^{ss}\left(1-\varepsilon^{-1}\right)}\equiv\psi_b\left(k^{ss}\right). \tag{43}$$

From (42)-(43), the simultaneous steady-state is characterized by the equilibrium condition $\psi_a\left(k^{ss}\right)=\psi_b\left(k^{ss}\right)$, and the existence and uniqueness of the equilibrium depends on the properties of $\psi_a\left(k^{ss}\right)$ and $\psi_b\left(k^{ss}\right)$. In this regard, function $\psi_a\left(k^{ss}\right)$ is decreasing and convex: from $\theta\geq 1$ and the Inada conditions, we have

$$\lim_{k^{ss}\to 0}=\infty,\quad \lim_{k^{ss}\to\infty}=0,\quad \psi_a'<0 \text{ for any } k^{ss}>0. \tag{44}$$

As regards $\psi_b\left(k^{ss}\right)$, the term $f\left(k^{ss}\right)-f_k^{ss}k^{ss}$ equals the marginal product of efficient labor, strictly increasing in $k\equiv K/\left(NB\right)$. As a consequence,

$$\lim_{k^{ss}\to 0}=0,\quad \lim_{k^{ss}\to\infty}=\infty,\quad \psi_b'>0 \text{ for any } k^{ss}>0. \tag{45}$$

Properties (44)-(45) imply that there always exists a unique $k^{ss}>0$ implying $\psi_a\left(k^{ss}\right)=\psi_b\left(k^{ss}\right)$, i.e. a symultaneous steady-state equilibrium $\left(c_{ss},k_{ss}\right)$ of system (13)-(14).

Proof of Lemma 1. From definition (12), the derivatives of $\Phi\left(k,c\right)$ read

$$\Phi_k\left(k,c\right)=-\frac{\theta-1}{\varepsilon\theta-1}\Phi\left(k,c\right)\frac{f_k\left(k\right)}{f\left(k\right)}<0 \text{ and } \Phi_c\left(k,c\right)=-\frac{1}{\varepsilon\theta-1}\Phi\left(k,c\right)\frac{1}{c}<0. \tag{46}$$

Also notice that, setting $\dot{c}=0$ and $\left(c,k\right)=\left(c^{ss},k^{ss}\right)$ in (14), the symultaneous steady-state is characterized by

$$f_k\left(k^{ss}\right)\cdot\Phi\left(k^{ss},c^{ss}\right)=\tilde{\rho}/\left(1-\varepsilon^{-1}\right). \tag{47}$$

The local stability properties of the symultaneous steady state can be studied by linearizing system (13)-(14) around $\left(c^{ss},k^{ss}\right)$. The relevant derivatives read

$$\frac{d\dot{k}}{dk}=f_k\left(k\right)\Phi\left(k,c\right)+f\left(k\right)\Phi_k\left(k,c\right)-\left(\tilde{\rho}-\rho\right)=f_k\left(k\right)\Phi\left(k,c\right)\left(1-\frac{\theta-1}{\varepsilon\theta-1}\right)-\left(\tilde{\rho}-\rho\right), \tag{48}$$

$$\frac{d\dot{k}}{dc}=f\left(k\right)\Phi_c\left(k,c\right)-1=-\left(1+\frac{f\left(k\right)}{c}\cdot\frac{\Phi\left(k,c\right)}{\varepsilon\theta-1}\right), \tag{49}$$

$$\frac{d\dot{c}}{dk}=\left(1-\varepsilon^{-1}\right)c\left[f_{kk}\left(k\right)\Phi\left(k,c\right)+f_k\left(k\right)\Phi_k\left(k,c\right)\right], \tag{50}$$

$$\frac{d\dot{c}}{dc}=f_k\left(k\right)\left(1-\varepsilon^{-1}\right)\left[\Phi_c\left(k,c\right)c+\Phi\left(k,c\right)\right]-\tilde{\rho}=f_k\left(k\right)\Phi\left(k,c\right)\left(1-\varepsilon^{-1}\right)\frac{\varepsilon\theta-2}{\varepsilon\theta-1}-\tilde{\rho}, \tag{51}$$

where we have used (46) to eliminate Φ_k and Φ_c in the various cases. Evaluating these derivatives at the steady-state (c^{ss}, k^{ss}), and using (47) to substitute $f_k(k)\Phi(k,c)$ in (48) and (51), we obtain

$$q_1 \equiv \left.\frac{d\dot{k}}{dk}\right|_{c^{ss},k^{ss}} = \rho + \frac{\tilde{\rho}}{\varepsilon\theta - 1} > 0 \tag{52}$$

$$q_2 \equiv \left.\frac{d\dot{k}}{dc}\right|_{c^{ss},k^{ss}} = -\left[1 + \frac{f(k^{ss})\tilde{\rho}}{c^{ss} f_k(k^{ss})(\varepsilon\theta - 1)(1 - \varepsilon^{-1})}\right] < 0 \tag{53}$$

$$q_3 \equiv \left.\frac{d\dot{c}}{dk}\right|_{c^{ss},k^{ss}} = c^{ss}\tilde{\rho}\left[\frac{f_{kk}(k^{ss})}{f_k(k^{ss})} - \frac{f_k(k^{ss})}{f(k^{ss})} \cdot \frac{\theta - 1}{\varepsilon\theta - 1}\right] < 0, \tag{54}$$

$$q_4 \equiv \left.\frac{d\dot{c}}{dc}\right|_{c^{ss},k^{ss}} = -\frac{\tilde{\rho}}{\varepsilon\theta - 1} < 0, \tag{55}$$

where we have used (46) to substitute $\Phi_k(k^{ss}, c^{ss})$ in (50), and (47) to eliminate all the remaining terms with $\Phi(k^{ss}, c^{ss})$. Expressions (52)-(55) are the coefficients in the Jacobian matrix of the linearized system

$$\begin{pmatrix} \dot{\overline{k - k^{ss}}} \\ \dot{\overline{c - c^{ss}}} \end{pmatrix} = \begin{pmatrix} q_1 & q_2 \\ q_3 & q_4 \end{pmatrix} \begin{pmatrix} k - k^{ss} \\ c - c^{ss} \end{pmatrix}, \tag{56}$$

and the eigenvalues of the Jacobian are computed as the roots (μ_1, μ_2) of the second-order equation

$$\mu^2 - (q_1 + q_4)\mu + [q_1 q_4 - q_2 q_3] = 0. \tag{57}$$

From (52) and (55) we have $q_1 + q_4 = \rho$, whereas (53)-(54) imply that the term inside the square brackets is strictly negative. As a consequence, the roots of (57) are real and of opposite sign,

$$\mu_{1,2} = (\rho/2) \pm (1/2)\sqrt{\rho^2 - 4[q_1 q_4 - q_2 q_3]},$$

which implies that the linearized system (56) is saddle-point stable. This implies that, for a given initial condition $k(0) = k_0 = K_0/(B_0 N_0)$, there exists a unique initial value $c(0) = \check{c}$ such that the system converges to (c^{ss}, k^{ss}). We now prove that the unique path associated with $c(0) = \check{c}$ is necessarily the optimal path because, as in the standard Ramsey model, all other paths diverging from (c^{ss}, k^{ss}) would violate either the accumulation constraint of the economy or the transversality condition (39).

Suppose that, for a given initial condition $k_0 = K_0/(B_0 N_0)$, the planner chooses $c(0) \neq \check{c}$. The initial level of normalized consumption does not lie on the stable arm of the saddle leading towards (c^{ss}, k^{ss}), and the resulting path of the economy is characterized by different asymptotic growth rates for normalized capital and normalized consumption:

$$c(0) \neq \check{c} \quad \Longrightarrow \quad \lim_{t\to\infty}(\dot{c}(t)/c(t)) \gtrless \lim_{t\to\infty}\left(\dot{k}(t)/k(t)\right),$$

which implies two classes of paths to consider. First, suppose that $c(0)$ is such that $\lim_{t\to\infty}(\dot{c}/c) > \lim_{t\to\infty}\left(\dot{k}/k\right)$. In this case, we have $\lim_{t\to\infty}(c/k) = +\infty$ and, since χ is bounded between 0 and 1 and $f(k)$ is strictly concave, the accumulation constraint (8) would imply $\lim_{t\to\infty}\left(\dot{k}/k\right) = -\infty$. However, this yields a violation of the non-negativity constraint $k(t) \geq 0$ from some finite t onwards, so that the class of paths with $\lim_{t\to\infty}(\dot{c}/c) > \lim_{t\to\infty}\left(\dot{k}/k\right)$ cannot be optimal. Second, suppose that $c(0)$ is such that $\lim_{t\to\infty}(\dot{c}/c) < \lim_{t\to\infty}\left(\dot{k}/k\right)$. In this case, $\lim_{t\to\infty}(c/k) = 0$, and the optimal asymptotic growth rate of the capital-consumption ratio would be[25]

$$\lim_{t\to\infty}\left(\frac{\dot{k}(t)}{k(t)} - \frac{\dot{c}(t)}{c(t)}\right) = \rho + \lim_{t\to\infty}\left\{\left[\frac{Y(t)}{K(t)} - F_K(t)\left(1-\varepsilon^{-1}\right)\right](1-\chi(t))\right\}. \quad (58)$$

Since $Y = F(K, BN)$ is linearly homogeneous, the term $(Y/K) - F_K$ equals $F_{BN}k^{-1}$, which is strictly positive, and this implies that the term in square brackets in (58) is strictly positive. Since $\chi(t)$ is bounded between 0 and 1, the term in curly brackets in (58) is necessarily non-negative, implying that an optimal path with $\lim_{t\to\infty}(\dot{c}/c) < \lim_{t\to\infty}\left(\dot{k}/k\right)$ would be characterized by

$$\lim_{t\to\infty}\left(\frac{\dot{k}(t)}{k(t)} - \frac{\dot{c}(t)}{c(t)}\right) \geq \rho. \quad (59)$$

However, this implies a violation of the transversality condition on capital: since $\dot{\lambda}^k/\lambda^k = -\dot{c}/c$, a necessary condition to satisfy (39) is $\lim_{t\to\infty}\left[\left(\dot{k}/k\right) - (\dot{c}/c)\right] < \rho$, which is obviously incompatible with (59). Hence, the class of paths with $\lim_{t\to\infty}(\dot{c}/c) < \lim_{t\to\infty}\left(\dot{k}/k\right)$ cannot be optimal. It follows from the above results that the unique path satisfying all the necessary conditions for optimality is the one characterized by the initial value $c(0) = \check{c}$ which yields convergence towards the symultaneous steady-state equilibrium (c^{ss}, k^{ss}). $\square$

Derivation of system (19)-(20). From (17), we have

$$c = (\Omega_0 B_0 \Gamma^\varepsilon)^{\frac{\varepsilon\theta-1}{\varepsilon}} \bar{p}^{-\frac{\varepsilon\theta-1}{\varepsilon}} f(k)^{1-\frac{1}{\varepsilon}}. \quad (60)$$

Plugging (60) in (13), the growth rate of $k(t)$ equals

$$g(k) = (f(k)/k)\,\Phi(k,c) - (\Omega_0 B_0 \Gamma^\varepsilon)^{\frac{\varepsilon\theta-1}{\varepsilon}} \bar{p}^{-\frac{\varepsilon\theta-1}{\varepsilon}} (f(k)/k)\, f(k)^{-\frac{1}{\varepsilon}} - (\tilde{\rho} - \rho),$$

[25]From (8) and (10), the growth rate of the capital-consumption ratio along an optimal path is

$$\frac{\dot{k}(t)}{k(t)} - \frac{\dot{c}(t)}{c(t)} = \left[\frac{f(k(t))}{k(t)} - f_k(k(t))\left(1-\varepsilon^{-1}\right)\right](1-\chi(t)) - \frac{c(t)}{k(t)} + \rho.$$

Setting $\lim_{t\to\infty}(c/k) = 0$, and recalling that $f(k)/k = Y/K$ and $f_k(k) = F_K$, we obtain expression (58).

where we can substitute $\Phi(k,c) = (\Omega_0 B_0)^{-\frac{1}{\varepsilon}} \bar{p}^{\frac{1}{\varepsilon}} f(k)^{-\frac{1}{\varepsilon}}$ from (17) to obtain

$$g(k) = (f(k)/k) f(k)^{-\frac{1}{\varepsilon}} (\Omega_0 B_0)^{-\frac{1}{\varepsilon}} \bar{p}^{\frac{1}{\varepsilon}} \left[1 - \left(\Omega_0^\theta B_0^\theta \Gamma^{\varepsilon\theta-1}\right) \bar{p}(t)^{-\theta}\right] - (\tilde{\rho} - \rho). \quad (61)$$

When the technology is Cobb-Douglas, $f(k) = k^\alpha$, we have $(f(k)/k) f(k)^{-\frac{1}{\varepsilon}} = k^{\alpha-1-\alpha/\varepsilon}$. Plugging this result in (61), and defining the constants $\varphi_2 \equiv (\Omega_0 B_0)^{-\frac{1}{\varepsilon}} > 0$, $\varphi_3 \equiv \Omega_0^\theta B_0^\theta \Gamma^{\varepsilon\theta-1} = \sigma/(\varsigma\varepsilon\theta) > 0$ and $\varphi_4 \equiv \tilde{\rho} - \rho = \delta + n + \pi > 0$, we obtain (20). As regards (19), re-write (14) as

$$g(c) = \alpha(f(k)/k) \Phi(k,c) \left(\frac{\varepsilon - 1}{\varepsilon}\right) + \frac{\varepsilon}{\varepsilon\theta - 1}\tilde{\rho}, \quad (62)$$

where we have used $f_k = \alpha(f(k)/k)$ and $g(c) \equiv \dot{c}/c$ for the consumption growth rate. Next time-differentiate (17) to get

$$g(\bar{p}) = \frac{\varepsilon - 1}{\varepsilon\theta - 1} g(f(k)) - \frac{\varepsilon}{\varepsilon\theta - 1} g(c), \quad (63)$$

where, given $f(k) = k^\alpha$, the growth rate of normalized output equals $g(f(k)) = \alpha g(k)$. Plugging $g(f(k)) = \alpha g(k)$ and substituting $g(k)$ with (61), and substituting $g(c)$ by means of (62), we obtain

$$g(\bar{p}) = \frac{\varepsilon\tilde{\rho} - \alpha(\tilde{\rho} - \rho)(\varepsilon - 1)}{\varepsilon\theta - 1} - \frac{\varepsilon - 1}{\varepsilon\theta - 1} \alpha(f(k)/k) \Phi(k,c) \left(\Omega_0^\theta B_0^\theta \Gamma^{\varepsilon\theta-1}\right) \bar{p}^{-\theta}. \quad (64)$$

Substituting $\Phi(k,c) = (\Omega_0 B_0)^{-\frac{1}{\varepsilon}} \bar{p}^{\frac{1}{\varepsilon}} f(k)^{-\frac{1}{\varepsilon}}$ from (17), and defining the constants $\varphi_0 \equiv \frac{\varepsilon\tilde{\rho} - \alpha(\tilde{\rho} - \rho)(\varepsilon - 1)}{\varepsilon\theta - 1} = \frac{\varepsilon\tilde{\rho}(1-\alpha) + \alpha\tilde{\rho} + \alpha\rho(\varepsilon - 1)}{\varepsilon\theta - 1} > 0$ and $\varphi_1 \equiv \alpha\frac{\varepsilon - 1}{\varepsilon\theta - 1} (\Omega_0 B_0)^{\theta - \frac{1}{\varepsilon}} \Gamma^{\varepsilon\theta-1} > 0$, we obtain (19).

Derivation of (23), (24) and (25). From, the derivatives of the right hand sides of (19)-(20) respectively exhibit the signs $\partial g(\bar{p})/\partial\bar{p} > 0$ and $\partial g(\bar{p})/\partial k > 0$, $\partial g(k)/\partial p > 0$ and $\partial g(k)/\partial k < 0$. Hence, the coefficient matrix of the linearized system is given by

$$\begin{array}{ll} m_1 \equiv \partial g(\bar{p})/\partial\bar{p}|_{\bar{p}^{ss}} > 0, & m_2 \equiv \partial g(\bar{p})/\partial k|_{\bar{p}^{ss}} > 0, \\ m_3 \equiv \partial g(k)/\partial\bar{p}|_{k^{ss}} > 0, & m_4 \equiv \partial g(k)/\partial k|_{k^{ss}} < 0, \end{array}$$

which proves (23)-(24). Given the above signs, system (23)-(24) displays two real roots of opposite signs, the stable root being

$$\bar{\mu} \equiv (1/2)\left[(m_1 + m_4) - \sqrt{(m_1 + m_4)^2 - 4(m_1 m_4 - m_2 m_3)}\right] < 0.$$

The stable arm equation is given by $\frac{k(t) - k_{ss}}{\bar{p}(t) - \bar{p}_{ss}} = \frac{\bar{\mu} - m_1}{m_2}$, where $\bar{\mu} < 0$, $m_1 > 0$, and $m_2 > 0$ imply that the right hand side is a strictly negative constant, $\phi \equiv \frac{\bar{\mu} - m_1}{m_2} < 0$.

Derivation of (26) and Proof of Proposition 2. Since output per capita equals $\bar{y}(t) = B(t)k(t)^{\alpha}$, its growth rate is given by $g(\bar{y}(t)) = \pi + \alpha g(k(t))$. Plugging (24) in this expression, we have

$$g(\bar{y}(t)) = \pi + \alpha\left[m_3(\bar{p}(t) - \bar{p}_{ss}) + m_4(k(t) - k_{ss})\right].$$

Eliminating $(k(t) - k_{ss})$ by means of the stable-arm equation (25) and rearranging terms yields

$$(\bar{p}(t) - \bar{p}_{ss}) = \frac{g(\bar{y}(t)) - \pi}{\alpha(m_3 + m_4\phi)}.$$

Plugging this expression in (23), and using (25) to eliminate $(k(t) - k_{ss})$, we obtain (26). Defining $H_1 \equiv \frac{m_1}{\alpha(m_3+m_4\phi)}$, $H_2 \equiv -\phi m_2$ and $H_0 \equiv H_2\bar{p}_{ss} - \pi H_1$, we obtain eqaution (26) in Proposition 2. Since $m_1 > 0$, $\alpha > 0$, $m_3 > 0$, $m_4 < 0$ and $\phi < 0$, coefficients H_1 and H_2 are both strictly positive, which completes the proof. □

A note on asymmetric rates of technical progress. In the main text - see the discussion below (5) - we have motivated our assumption of symmetric rates of labor-augmenting and emission-reducing progress (i.e. $\dot{\Omega}/\Omega = -\dot{B}/B = -\pi$), by claiming that parameters must satisfy a critical condition for obtaining bounded levels of pollution per capita in the long run. We now show that, given the aim of the present analysis, this assumption is innocuous. Consider a more general model in which a generic rate of emission-reducing technical progress implies $\Omega(t) = \Omega_0 e^{-\omega t}$, with $\omega > 0$. From (3) and (1), the dynamics of pollution per capita, $\bar{p}(t) = p(t)B(t)$, are given by

$$\bar{p}(t) = \Omega(t)\Phi(k(t), c(t))^{\varepsilon} f(k(t))B(t) = \Omega_0 B_0 e^{-(\omega-\pi)t}\Phi(k(t), c(t))^{\varepsilon} f(k(t)), \tag{65}$$

where the optimal path of $1 - \chi(t)$ is given by an optimality condition that is slightly different from (11) - see equation (41) above. The crucial point is that $\chi(t)$ is bounded between zero and unity, independently of the characteristics of the optimality condition determining the optimal path of ICT effort. This implies the following Lemma, the proof of which is left to a footnote[26]

Lemma 3 *In any optimal path characterized by* $\lim_{t\to\infty} k(t) = \tilde{k}$, *assuming* $\omega > \pi$ *implies*

$$\lim_{t\to\infty} \bar{p}(t) = 0.$$

[26]Lemma 3 is proved as follows. Taking the limit in (65), we obtain

$$\lim_{t\to\infty} \bar{p}(t) = \Omega_0 B_0 \cdot \lim_{t\to\infty}\left[f(k(t))(1-\chi(t))^{\varepsilon}\right]e^{-(\omega-\pi)t}.$$

Since $\lim_{t\to\infty} k(t) = \tilde{k}$, the limit of the term in square brackets, $\lim_{t\to\infty} f(k)(1-\chi)^{\varepsilon}$, is bounded from above by $f\left(\tilde{k}\right) < \infty$. When $\omega > \pi$, the exponential term in the above expression yields $\lim_{t\to\infty} e^{-(\omega-\pi)t} = 1/\infty$, which implies $\lim_{t\to\infty} \bar{p}(t) = 0$. □

Lemma 3 establishes that any optimal path along which normalized capital converges to a finite steady state (which is always the case in neoclassical models of balanced growth) is characterized by zero pollution per capita in the long run. On the one hand, this result provides an *ex-post* proof of our claim that the relevant 'sustainability condition' in our model is $\dot{\Omega}/\Omega \leq -\pi$.[27] On the other hand, Lemma 3 implies that focusing, as we did, on the polar case $\omega = \pi$ is innocuous for the problem under study. Indeed, the general aim of our theoretical analysis is to provide a micro-foundation of the convergence equation (29), and since $\omega > \pi$ implies bounded pollution per capita in the long run (with the additional restriction that $\bar{p}$ approaches zero asymptotically), the main prediction of the model does not change: as the economy approaches balanced growth paths, pollution per capita displays convergence towards a finite limit. The only relevant modification induced by a generic sustainable rate of emission-reducing progress (i.e. $\Omega(t) = \Omega_0 e^{-\omega t}$ with $\omega \geq \pi$) is a technical complication: if $\Omega(t) B(t)$ is not constant, the optimal level of ICT effort becomes a function of output, consumption, *and time* (the term in square brackets in (41) depends explicitly on $e^{-(\omega-\pi)t}$), and this implies that we cannot reduce the dynamic analysis to a two-by-two autonomous system like (13)-(14). The optimal path may still be studied by assessing the stability properties of a three-by-three dynamic system in the variables $(c(t), k(t), \chi(t))$, but this is clearly not a central issue for the present analysis.

[27]Combining Lemma 3 with result (18), it follows that pollution per capita is bounded in the long run if the rate of emission-reducing technical progress, ω, is either equal or strictly greater than the rate of of labor-augmenting technical progress.

References

Andreoni, J. and A. Levinson, "The simple analytics of the environmental Kuznets curve," *Journal of Public Economics*, 2001, *80*, 269–286.

Barro, R.J. and X. Sala i Martin, *Economic Growth*, second ed., Massachusetts: The MIT Press, 2004.

Bratberg, E., S. Tjotta, and T. Oines, "Do voluntary environmental agreements work?," *Journal of Environmental Economics and Managment*, 2005, *50*, 583–597.

Breusch, T.S. and A.R. Pagan, "A simple test for heteroscedasticity and random coefficient variation," *Econometrica*, 1979, *47*, 1287–1294.

Brock, W.A. and M.S. Taylor, "Economic growth and the environment: a review of theory and empirics," Working paper 10854, NBER Oct. 2004.

__ and __ , "The Green Solow Model," Working paper 10557, NBER June 2004.

Bulte, E., J. List, and M.C. Strazicich, "Regulatory Federalism and the Distribution of Air Pollutant Emissions," *Journal of Regional Science*, 2007, *47(1)*, 155–178.

Fan, J. and R. Li, "New estimation and model selection procedure for semiparametric modeling in longitudinal data analysis," *Journal of American Statistical Association*, 2004, *99*, 710–723.

__ , W. Haerdle, and E. Mammen, "Direct Estimation of Additive and Linear Components for High-Dimensional Data," *The Annals of Statistics*, 1998, *26*, 943–971.

Finus, M. and S. Tjotta, "The Oslo Protocol on sulfur reduction: the great leap forward?," *Journal of Public Economics*, 2003, *87*, 2031–2048.

Greene, W.H., *Econometric Analysis*, 5th ed., Prentice Hall International, 2003.

Gu, C., *Smoothing Spline ANOVA Models*, New York: Springer, 2002.

Hastie, T. and R. Tibshirani, *Generalized Additive Models*, Chapman and Hall, 1990.

Hsiao, C., Q. Li, and J.S. Racine, "A Consistent Model Specification Test With Mixed Categorical and Continuous Data," *Journal of Econometrics*, 2007.

Johnson, P.A., "A Nonparametric Analysis of Income Convergence accross the U.S. States," *Economics Letters*, 2000, *69*, 219–223.

Koenker, R., "A Note on Studentizing a Test for Heteroscedasticity," *Journal of Econometrics*, 1981, *17(1)*, 107–112.

Li, Q. and J.S. Racine, "Nonparametric estimation of distributions with categorical and continuous data," *Journal of Multivariate Analysis*, 2003, *86*, 266–292.

List, J.A., "Have Air Pollutant Emissions Converged Among U.S. Regions? Evidence from Unit Root Tests," *Southern Economic Journal*, 1999, *66* (1), 144–155.

Maasoumi, E., J. Racine, and T. Stengos, "Growth and convergence: A profile of distribution dynamics and mobility," *Journal of Econometrics*, 2007, *136*, 483–508.

Maddison, A., "Historical Statistics," Online data, Groningen Growth and

Development Center, University of Groningen, October 2008. http://www.ggdc.net/maddison.

Murdoch, J.C., T. Sandler, and K. Sargent, "A Tale of Two Collectives: Sulphur versus Nitrogen Oxides Emission Reduction in Europe," *Economica*, 1997, *64*, 281–301.

Ordás Criado, C. and J.-M. Grether, "Convergence in CO_2 per capita emissions: a robust distributional approach.," Working paper, 17th Annual Conference of the European Association of Environmental and Resource Economists (EAERE), Amsterdam 2009.

Quah, D., "Empirical cross-sectional dynamics in economic growth," *European Economic Review*, 1993, *37*, 426–434.

Racine, J. and Q. Li, "Nonparametric estimation of regression functions with both categorical and continuous data," *Journal of Econometrics*, 2004, *119*, 99–130.

Silverman, B.W., *Density estimation for statistics and data analysis*, London: Chapman and Hall, 1986.

Stockey, N., "Are there limits to growth," *International Economic Review*, 1998, *39(1)*, 1–31.

Stone, C., "Additive regression and other nonparametric models," *Annals of Statistics*, 1985, *13*, 689–705.

Strazicich, M.C. and J.A. List, "Are CO_2 Emissions Levels Converging Among Industrial Countries?," *Environmental and Resource Economics*, 2003, *24*, 263–271.

Van der Ploeg, F. and C. Withagen, "Pollution control and the Ramsey problem," *Environmental and Resource Economics*, 1991, *1(2)*, 215–236.

Van, P. Nguyen, "Distribution Dynamics of CO_2 Emissions," *Environmental and Resource Economics*, 2005, *32*.

Wood, S., "Modeling and Smoothing parameter estimation with multiple quadratic penalties," *Journal of Royal Statistical Aociety*, 2000, *Series B 65*, 95–114.

—, "Stable and efficient multiple smoothing parameter estimation for generalized additive models," *Journal of American Statistical Association*, 2004, *99*, 673–686.

—, *Generalized Additive Models: An Introduction with R*, Chapman & Hall, Texts in Statistical Sciences, 2006.