National Carbon Dioxide Emissions:

Geography Matters

Abstract

This article examines the role of geographical factors as determinants of cross-country differences in per capita carbon dioxide emissions. Such differences have been explained by economists mostly in terms of per capita income. Geographical factors on the other hand have been neglected by economic analysis. We examine the effects of cold and hot climates, transportation requirements and the availability of renewable energy sources on emissions. We find that with the exception of cooling requirements as measured by hot climates, all these geographical factors are statistically significant determinants of emissions in accordance with our expectation. Furthermore, cold climates and the availability of renewable resources are also substantively important.

Key words: Global, carbon dioxide, quantitative analysis, climate, transportation, renewable energy

Introduction

Economists have analysed the determinants of differences in carbon dioxide (CO_2) emissions across countries for more than a decade (see, for example, Shafik 1994; Grossman and Krueger 1995; Holtz-Eakin and Selden 1995; Schmalensee et al. 1998; Galeotti and Lanza 1999; Ravallion, Heil and Jalan 2000; Heil and Selden 2001). They mainly explain such differences with the help of per capita income as the independent variable in an analytical framework called the Environmental Kuznets Curve (EKC). In this framework, emissions first rise with increasing income, but at a decreasing rate. For some pollutants, a turning point is within the range of existing income levels such that emissions are predicted to fall in some countries with high income levels (Cole and Neumayer 2002). In the case of CO_2 emissions, however, the estimated turning point is typically much beyond even the highest existing income level such that emissions merely rise at a decreasing rate with higher income levels, but are not predicted to fall in any country.

Economic analysis has not been interested in the impact of geographical factors on such emissions. Geography should clearly matter, however. For example, we would expect countries faced with cold winter months to have higher heating requirements than others with very mild or even warm winters. Similarly, countries with hot summers might have greater cooling requirements with consequently greater emissions. We would expect countries which nature endowed with the gift of renewable energy resources to have lower emissions.[1] Finally, we would expect countries with populations spatially scattered across a big land area to have higher transportation requirements and therefore higher emissions than small countries or those with highly concentrated clusters of population.

Economic analysis has not completely overlooked the potential importance of geography, since many studies include country-specific fixed effects or at least model these as random effects forming part of the stochastic error term. Sometimes the inclusion of country-specific fixed effects is explicitly justified with reference to ‘climatic patterns or resource endowments’ as in Heil and Selden (2001, p. 38). However, the inclusion of such effects is unsatisfactory for mainly two reasons. First, it meshes geographical factors together with any other country-specific fixed effects. Second, in failing to include geographical variables explicitly in the estimations, these studies can offer no insights into the distinct impact of different geographical factors on CO_2 emissions.

This article attempts to demonstrate that geography indeed matters when it comes to CO_2 emissions. It improves upon preliminary estimates undertaken within a more basic estimation framework (Neumayer 2002). We will show that cold climates, limited access to renewable energy sources and higher transportation requirements are all statistically significantly associated with higher CO_2 emissions. In contrast, cooling requirements as approximated by hot climates exert no statistically significant impact upon emissions.

Somewhat surprisingly, geographical aspects have not played much of a role in the negotiations of emission reductions in Kyoto, Japan, in 1997 (Jones 2001). They have, however, at times been considered in debates about what a fair allocation of emission reduction obligations would look like. For example, Grubb et al. (1992, p. 314) examine, without endorsment, “reasonable emissions” as one criterion for an internationally just allocation rule. They define a ‘reasonable level of emissions for each country’ as the ‘level that would support a consistent, modest standard of living, *given the national climatic and other conditions* [emphasis added]. Permits would be granted

for emissions at this level, but not for those “luxury” emissions in excess of this amount’. The Intergovernmental Panel on Climate Change (IPCC 1995, p. 104) contemplates a similar allocation rule under the heading “basic needs”. Such a rule would allow countries ‘the right to emit the minimum levels of greenhouse gases needed to meet the basic needs of their citizens (…). It would perhaps be close to the allocation of emission permits according to population, although basic needs could vary from country to country *depending on climate and other matters* [emphasis added]’. As a final example, consider the attempt by Benestad (1994) to construct a formula for just allocation of CO_2 emission rights according to energy needs, including such things as a country’s heating and cooling requirements, transportation needs, as well as renewable energy sources potential. Since this study examines the relative importance of a number of geographical factors explaining cross-country differences in CO_2 emissions, it can also shed some light on the relevance of these and other normative allocation rules that refer to such factors.

Research design

We use a panel of per capita CO_2 emissions covering the period 1960 to 1999 with up to 163 countries. The country coverage is entirely driven by the availability of data. We estimate variants of the following model:

$$\ln(E_{it}) = \beta_0 + \beta_1 \ln(Y_{it}) + \beta_2(\ln Y_{it})^2 + \beta_3 C_i + \beta_4 R_i + \beta_5 A_i + T_t + e_{it}$$

where countries are indicated by i and years by t. The variable *E* stands for per capita CO_2 emissions, *Y* is income per capita, *C* is a climate variable, *R* is the percentage of total energy consumption derived from renewable energy sources, *A* is a measure for

transportation requirements, T_t are (t-1) year-specific dummy variables and e_{it} is a stochastic error term. The dependent variable is logged to make its distribution less skewed, which typically makes the estimated model more compatible with distributional assumptions of the estimation (Wooldridge 2000). The logging of the income variable is common in the EKC literature, partly because it allows an elasticity interpretation of the estimated coefficients and partly because empirically the model fit is better with the income variable logged. Our main results would be similar if these variables were not logged. The year-specific dummy variables capture exogenous advances in carbon saving technology available to all countries.

For panel data, common estimators used are the fixed-effects and either the generalised least squares or the full maximum-likelihood random-effects estimators. The fixed-effects estimator cannot be used here since some of our variables do not vary over time. Instead of the random-effects estimators we use a generalised estimating equations (GEE) estimator that is asymptotically equivalent to the random-effects estimators. In addition, this estimator makes it very easy to compute standard errors that are robust to heteroscedasticity as well as so-called clustering, which means that observations are assumed to be independent only across countries, but not necessarily across time within any one country.[2] Zorn (2001, p. 470) calls for the use of GEE estimators for cases 'in which the standard assumption that the data are conditionally independent can be called into question'. Since observations within any one country are certainly not independent over time, using standard errors that are robust towards clustering by countries is clearly advantageous.

Total per capita CO_2 emissions from fossil fuel burning and cement manufacturing were taken from data compiled by the Carbon Dioxide Information Center, the definitive source for such data (Marland et al. 2002). Income is measured as real per

capita GDP in purchasing power parity taken from the Penn World Table 6.1 (Heston, Summers and Aten 2003). For our climate variables we use two proxy variables for cold climates and the consequent heating requirements. One is the average minimum temperature in degrees Centigrade in the coldest climatic season of the year. For countries located in the Northern hemisphere this means December, January and February, for those in the Southern hemisphere the temperatures refer to June, July and August. The other variable is the annual number of frost days, which is defined as the number of days in which daily minimum temperature drops below zero degrees Centigrade. Both variables are very highly correlated as one would expect. Neumayer (2002) suggests that the maximum temperature as a proxy for cooling requirements is not consistently and robustly associated with higher emissions. Nevertheless, we use the average maximum temperature in the hottest climatic seasons of the year as a further variable in extended estimations to our main analysis. All variables are taken from the climate data set for political areas described in Mitchell, Hulme and New (2002).

The share of total energy consumption derived from renewable resources is calculated from data in WRI (2003). As this is the variable with the lowest availability, our estimations are run once with and once without it. In addition to hydroelectricity, renewable resources also cover energy from primary solid biomass, thermal solar, photovoltaic solar, wind, biogas, liquid biomass, and tide, wave, and ocean. Fuel and waste renewable energy sources in the form of biomass are much used by poor developing countries. While they partly create CO_2 (and other greenhouse gas) emissions, they are usually not included in CO_2 emission data, which derive exclusively from estimates of fossil fuel burning and cement manufacturing. In as much as fuel and waste renewable energy sources substitute for fossil fuels, which would have otherwise been used, their consumption should lead to lower CO_2 emissions thus measured. In this

respect, they do not differ from other substitute renewable energy sources that entail few CO_2 emissions, such as hydroelectricity. It is therefore correct to include them for the purposes of explaining cross-country CO_2 emissions here.

As concerns transportation requirements, big countries have higher transportation requirements as goods and people are typically moved over longer distances. However, it would be highly misleading to simply take a country’s total land area as a proxy for its transportation requirements. This is because often huge parts of big countries such as Canada or the Russian Federation are sparsely inhabited, if at all. To measure transportation requirements, we take two proxy variables. CIESIN (2001) provides geographical information systems data from some time in the late 1990s on the percentage of total land area that is either urbanised (as indicated by lights at night) or used for agriculture. This provides a good proxy to the share of total land area impacted by human activities. The proxy for a country’s transportation requirements is then the share of total land area impacted by human activities (data for land area taken from World Bank 2002). As an alternative proxy variable, we use the total length of the road network, both paved and unpaved. Poor over-time availability made it necessary to take the average length of the road network in the 1990s, with data taken from IRF (various year). Roads are built where people live. Big countries will have a larger road network, but its length also depends on how scattered the population is over the entire land area. For example, Australia has a small total length of road network relative to its land area as its population tends to be clustered in a few population centres along the coast line. India has a smaller land area, but much larger total length of road network as its population is scattered almost all over the country. Our two proxy variables are very highly correlated as one would hope , given that they are supposed to approximate the

same underlying concept transportation requirements. Table 1 provides summary statistics on all variables, Table 2 a matrix of bivariate correlation coefficients.

< Insert tables 1 and 2 about here >

Results

We start with the model that excludes the renewable energy variable and therefore has the largest sample size. Column I of Table 3 shows the expected EKC results as the coefficient of the linear income term is positive, the coefficient of the squared term is negative and both are statistically significant. A higher minimum temperature during the cold season is associated with lower CO_2 emissions, a longer total road network with higher emissions, all in line with expectations. In column II we replace the temperature variable with the annual number of frost days, which is highly significant with the expected positive sign. In columns III and IV we repeat the last two estimations, but replace the total road length with the variable measuring the land area impacted by human beings. In both estimations, this variable has the expected positive sign, but it is marginally insignificant in column III and only marginally significant at the .1 level in column IV.

< Insert table 3 about here >

In Table 4 we similarly estimate four different models analogous to table 3, but add the variable measuring the share of total energy consumption derived from renewable resources. The statistical significance of the existing variables is often slightly reduced

in this smaller sample, but all remain significant at the .05 level with the expected signs. The share of renewable resources is negatively associated with emissions and highly statistically significant. The variable measuring the land area impacted by human beings is now significant in both model specifications at the higher confidence level of .05.

< Insert table 4 about here >

We will now include the maximum temperature during the hot climatic season as an additional control variable to our preferred model. Our preferred model is that in which heating requirements are approximated by the annual number of frost days and transportation requirements by the total length of road network. This model is preferred because these variables are more clearly statistically significant than their respective alternatives. We find the maximum temperature variable to be highly insignificant (column I of Table 5). This holds true as well for the smaller sample with the renewable resource variable included (column II). Indeed, the same holds true for any other model specification (detailed results not reported).

As further sensitivity analysis, we tested whether our results are due to the influence of outliers. Belsley, Kuh and Welsch (1980) suggest excluding observations as outliers that have both high residuals and a high leverage. Their criterion is to exclude an observation if its so-called DFITS is greater than twice the square root of (k/n), where k is the number of independent variables and n the number of observations. DFITS is defined as the square root of ($h_i/(1-h_i)$), where h_i is an observation's leverage, multiplied by its studentized residual. Applying this criterion leads to the exclusion of 4 countries and 257 observations in the estimation with the larger sample size and the exclusion of 2 countries and 122 observations in the smaller sample where the

renewable resource variable is included. Re-estimating our preferred model with the remaining observations leads to the results reported in columns III and IV of Table 5. Clearly, the results from our main analysis are not driven by the presence of outliers.

< Insert table 5 about here >

Discussion

In accordance with the existing literature we find a non-linear effect of per capita income levels on per capita CO_2 emissions. Theoretically, therefore, there exists a level of income beyond which emissions are predicted to decrease with further increases in income. This so-called turning point can be calculated as $-\beta_1/(2\beta_2)$, where β_1 is the coefficient of the linear and β_2 the coefficient of the squared income term. The turning point in our estimations lies between \$55000 as a low estimate (based on results from table 3) and about \$90000 as a high estimate (based on results from Table 4). In accordance with the existing literature, we therefore find a turning point that is beyond any currently existing per capita income levels such that CO_2 emissions in all countries are predicted to increase with higher income levels, albeit at a decreasing rate.

What about our geographical variables? To get a feeling for the importance of these variables, it is a good idea to see how much predicted emissions change due to a substantial increase in the variable, where we use a one standard deviation increase to mean substantial. A one standard deviation increase in the average minimum temperature in the cold season reduces CO_2 emissions by between 15 per cent (Table 4) and 41 per cent (Table 3). A one standard deviation increase in the annual number of frost days increases emissions by between 22 per cent (table 4) and 71 per cent (Table

3). These are clearly non-negligible differences in emissions, demonstrating that climatic factors are not only statistically significant but also substantively important.

In comparison, transportation requirements are less substantively important. The emission increase due to a one standard deviation increase in the total road length is estimated to be between about 8 per cent (Table 4) and 17 per cent (Table 3). The respective figures for the total land area impacted by human beings are 6 and 9 per cent. These emission increases are clearly more modest compared to the ones for climatic factors.

With respect to the availability of renewable energy sources, a one standard deviation increase in the share of total energy consumption satisfied with renewable sources is predicted to lower CO_2 emissions by about 42 per cent. Again, the availability of renewable energy sources has an impact on CO_2 emissions that is both statistically significant and substantively important.

Another way to gauge the importance of geographical factors is to compare the CO_2 emissions of two fictitious countries: one at the mean of all independent variables, the other with mean income levels, but geographically disadvantaged in the sense that, for example, the annual number of frost days and the total length of road network are one standard deviation above the mean, whereas the renewable resource share is one standard deviation below the mean. The difference in emissions amounts to .83 tons per capita where our calculations refer to the estimations in column II of Table 4. Given that mean CO_2 emissions per capita are .91 tons with a standard deviation of 1.32 tons, this is clearly a non-negligible difference in emissions.

What might explain the insignificance of the maximum temperature variable as a proxy for cooling requirements due to hot climates? One explanation could be that it is a bad proxy variable for cooling requirements. However, there is no strong reason why

this should be the case. A perhaps more convincing explanation could be that whereas heating represents a necessity good in cold climates with consumers having few alternatives if they do not want to freeze to death, cooling is likely to be a luxury good in hot climates. Those who can afford will have air conditioning and other cooling devices, those who cannot will not. Many countries with very hot climates are also relatively poor countries with many people not willing or able to afford air conditioning.

On the whole, we have demonstrated that geography matters when it comes to explaining variation in CO_2 emissions. Cold climates and the availability of renewable energy sources exert a statistically significant impact upon such emissions that is also substantively important. We found transportation requirements to be statistically significant as well, but less substantively important.

Our findings become policy relevant when it comes to debates over a fair allocation of emission rights. The results presented above give geographically disadvantaged countries some arguments at hand to request higher emission rights than are given to geographically advantaged countries. The world has only started to embark upon negotiating national emission rights. As emission reduction obligations become tougher in follow-up agreements to the Kyoto Protocol, with more countries required to reduce emissions, we can expect that countries pay much more attention to geographical factors than they have done so far. Since geography matters for CO_2 emissions it will also eventually matter for negotiations about emission reductions.

Acknowledgements

I would like to thank Francesca Pozzi from the Center for International Earth Science Information Network (CIESIN), Columbia University, for providing data on one of the

variables. I would also like to thank Timothy Mitchell for drawing my attention to their climate data set for political areas.

Notes

[1] Of course, how a country makes use of its natural endowment of renewable resources is to some extent also determined by political and economic decisions. But a fundamental dependence on geography exists. Hydroelectricity, for example, can only be generated if sufficiently voluminous water flows are available.

[2] At least within Stata, the statistical package used here, there is no easy option to use the random-effects estimators with such robust standard errors.

References

Belsley D A, Kuh E and Welsch R E 1980 *Regression diagnostics* Wiley, New York

Benestad O 1994 Energy needs and CO_2 emissions – constructing a formula for just distributions *Energy Policy* 22 725-734

CIESIN 2001 *Data on land area impacted by human activities as a percentage of total land area*. Center for International Earth Science Information Network, Columbia University, New York

Cole M A and Neumayer E 2002 *Economic growth and the environment in developing countries: what are the implications of the Environmental Kuznets Curve?* Working Paper, University of Birmingham and London School of Economics

Galeotti M and Lanza A 1999 Richer and cleaner? A study on carbon dioxide emissions in developing countries *Energy Policy* 27 565-573

Grossman G M and Krueger A B 1995 Economic growth and the environment *Quarterly Journal of Economics* 110 353-377

Grubb M, Sebenius J, Magalhaes A and Subak S 1992 Sharing the burden in **Mintzer I M** ed *Confronting climate change: risks, implications and responses* Cambridge University Press, Cambridge 305-322.

Heil M T and Selden T M 2001 International trade intensity and carbon emissions: a cross-country econometric analysis *Journal of Environment & Development* 10 35-49

Heston A, Summers R and Aten B 2003 *Penn World Table Version 6.1* Center for International Comparisons at the University of Pennsylvania (CICUP)

Holtz-Eakin D and Selden T M 1995 Stoking the fires? CO_2 emissions and economic growth *Journal of Public Economics* 57 85-101

IPCC 1995 *Climate change 1995 – economic and social dimensions of climate change* Oxford University Press, Oxford

IRF various years *World road statistics* International Road Federation, Geneva

Jones N 2001 Big chilly countries lose out under Kyoto *New Scientist* 2319 19

Marland G, Boden T and Andres R J 2002 *Global, regional and national fossil fuel CO_2 emissions: 1751-1999* Oak Ridge National Laboratory, Carbon Dioxide Information Center

Mitchell T D, Hulme M and New M 2002 Climate data for political areas *Area* 34 109-112

Neumayer E 2002 Can natural factors explain any cross-country differences in carbon dioxide emissions? *Energy Policy* 30 7-12

Ravallion M, Heil M and Jalan J 2000 Carbon emissions and income inequality *Oxford Economic Papers* 52 651-669

Schmalensee R, Stoker, T M and Judson, R A 1998 World carbon dioxide emissions: 1950-2050 *Review of Economics and Statistics* 80 15-27

Shafik, N 1994 Economic development and environmental quality: an econometric analysis *Oxford Economic Papers* 46 757-773

Wooldridge J M 2000 *Introductory econometrics: a modern approach* South-Western College Publishing, Cincinnati

World Bank 2002 *World development indicators* World Bank, Washington DC.

WRI 2003 *Earthtrends – the environmental information portal* (www.wri.org)

Zorn C J W 2001 Generalised estimation equation models for correlated data: a review with applications *American Journal of Political Science* 45 470-490

Table 1 Summary statistics

	N	Mean	Std. Dev.	Min.	Max.
CO_2 p.c. (metric tons)	4688	0.91	1.32	0.01	16.88
GDP p.c. (US$ of 1996)	4688	5982	6020	281.3	41354
minimum temperature (degree C)	4688	9.7	10.6	-28.1	24.2
maximum temperature (degree C)	4688	29.1	5.7	10.6	43.4
number of frost days (per annum)	4688	39.2	57.5	0	252.5
total length road network (km)	4688	210371	682341	246	6330325
total land area impacted by humans (sq km)	4688	197936	480399	63	2930918
renewable energy (% of energy consumption)	2881	0.31	0.31	0	1

Table 2 **Matrix of bivariate correlation coefficients**

	ln (CO_2 p.c.)	ln (GDP p.c.)	minimum temp.	maximum temp.	frost days	length road network	land area impacted
ln (GDP p.c.)	.8972						
minimum temperature	-.5530	-.5295					
maximum temperature	-.5051	-.5788	.7134				
frost days	.5407	.5349	-.9269	-.8338			
length road network	.2034	.1755	-.2217	-.0351	.2060		
land area impacted	.1026	.0234	-.2264	-.0096	.2004	.8282	
renewable energy share	-.8399	-.7462	.5179	.3058	-.4028	-.1457	-.0657

Note: N = 2881.

Table 3 Results with renewable energy share excluded

	I	II	III	IV
ln (GDP p.c.)	2.72***	2.72***	2.71***	2.71***
	(3.99)	(3.98)	(3.98)	(3.97)
[ln (GDP p.c.)]^2	-.12***	-.12***	-.12***	-.12***
	(3.04)	(3.05)	(3.03)	(3.04)
minimum temperature	-.05***		-.05***	
	(6.83)		(6.53)	
frost days		.01***		.01***
		(6.70)		(6.45)
length road network	2.22e-07***	2.26e-07***		
	(3.41)	(3.07)		
land area impacted			1.63e-07	1.80e-07*
			(1.51)	(1.71)
# countries	163	163	163	163
# observations	4688	4688	4688	4688

Note: Dependent variable is ln (CO_2 p.c.). General Estimation Equations (GEE) estimation. Absolute z-values in parentheses. Robust standard errors allow observations to be clustered by countries. Coefficients of constant and year-specific time dummies not shown. * statistically significant at .1 level ** at .05 level *** at .01 level.

Table 4 **Results with renewable energy share included**

	I	II	III	IV
ln (GDP p.c.)	2.50***	2.48***	2.48***	2.46***
	(3.01)	(3.02)	(2.99)	(3.00)
[ln (GDP p.c.)]^2	-.11**	-.11**	-.11**	-.11**
	(2.32)	(2.33)	(2.30)	(2.30)
renewable energy share	-1.77***	-1.79***	-1.78***	-1.80***
	(6.03)	(6.30)	(6.01)	(6.29)
minimum temperature	-.02***		-.02**	
	(2.60)		(2.41)	
frost days		.003***		.003***
		(3.33)		(3.15)
length road network	1.23e-07***	1.12e-07***		
	(3.81)	(3.84)		
land area impacted			1.38e-07**	1.20e-07**
			(1.98)	(1.94)
# countries	119	119	119	119
# observations	2881	2881	2881	2881

Note: Dependent variable is ln (CO_2 p.c.). General Estimation Equations (GEE) estimation. Absolute z-values in parentheses. Robust standard errors allow observations to be clustered by countries. Coefficients of constant and year-specific time dummies not shown. ** statistically significant at .05 level *** at .01 level.

Table 5 Sensitivity analysis

	I	II	III	IV
ln (GDP p.c.)	2.71***	2.51***	3.32***	2.12***
	(3.96)	(3.03)	(5.66)	(3.37)
[ln (GDP p.c.)]^2	-.12***	-.11**	-.15***	-.09**
	(3.03)	(2.34)	(4.29)	(2.47)
renewable energy share		-1.75***		-1.84***
		(6.06)		(7.35)
frost days	.012***	.005***	.009***	.003***
	(6.30)	(3.45)	(7.69)	(3.94)
length road network	2.05e-07***	9.85e-08***	1.86e-07***	1.02e-07***
	(2.58)	(2.93)	(3.14)	(4.12)
maximum temperature	.03	.02		
	(1.23)	(1.14)		
# countries	163	119	159	117
# observations	4688	2881	4431	2759

Note: Dependent variable is ln (CO_2 p.c.). General Estimation Equations (GEE) estimation. Absolute z-values in parentheses. Robust standard errors allow observations to be clustered by countries. Coefficients of constant and year-specific time dummies not shown. ** statistically significant at .05 level *** at .01 level.