Panel Data analysis of CO_2 emissions
Dany Yatim

February 27, 2017

1 The data

The panel analysis is done using data about 150 countries over the period 1990-2012. The study make use of data about CO2 emmissions (metric tons per capita), GDP (per capita), renewable energy consumption (share of total final energy consumption per year) and urban population (peoples per year) which are all obtained from data.worldbank.org (see Annex 1 and Annex 2 for the country list).

Table 1: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
co2	4.119	4.797	0.017	36.817	3402
gdp	8565.185	14188.099	64.81	113240	3325
re	39.642	33.338	0.001	100	3337
up	139.806	503.68	1.719	7524.7	3426

2 The model

In a try to catch the Kuznets curve's effect between CO2 and income highlighted by Gene M. Grossman and Alan B. Krueger in their study *Economic Growth and the Environment*, we use the GDP and GDP^2 as regressors. Hence, in order to underline a little more the improvement phase after the deterioration phase of the environmental state due to economic growth, we add the renewable energy (RE) consumption as regressor.
In a paper, Inmaculada Martínez-Zarzoso studied the impact of urbanization on CO2 emissions, and found a significant link between the later and urban population. In a try to replicate this result, we add urban population (UP) as regressor.

We make use of two models. The first contains only the later regressors with the first lag of CO2 in order to take in account its persistence in time, whereas the second one includes also the post-subprimes crisis period of 2009 and 2010 (timme dummies), in order to catch the crisis' effect on CO2 emissions.

Model 1 :

$$CO2_t = \alpha + \alpha_1 CO2_{t-1} + \beta_1 GDP + \beta_2 GDP^2 + \beta_3 RE + \beta_4 UP + \epsilon_t$$

Model 2 :

$$CO2_t = \alpha + \alpha_1 CO2_{t-1} + \beta_1 GDP + \beta_2 GDP^2 + \beta_3 RE + \beta_4 UP + \beta_5 yr2009 + \beta_6 yr2010 + \epsilon_t$$

Thus, it is chosen here to use a dynamic model. We make the assumption that the level of CO2 in a year is hightly determined by the past years' level of CO2. This is in accordance with C. Varotsos, M.-N. Assimakopoulos, and M. Efstathiou conclusions who showed that "*the fluctuations of the CO2 concentrations exhibit strong long-range persistence (almost 1/f – type), which signifies that the fluctuations in CO2 concentrations, from small time intervals to larger ones (up to 11 years) are positively correlated*".

We could make use of the consistent Anderson-Hsiao estimator but, in order to use all the informations available in the time dimension of the dataset, and thus improve the efficiency of the estimator, we jump to the Arellano-Bond estimator. Several reasons lead us to use instrumental variable estimators :

- Because the data concern 150 countries, the study observes the effect of the regressors for countries that differ in terms of developement and others unobserved heterogenities. In order to get rid of the constant term α and the individual effect, we take the first difference of the models presented before, which brings us to use instrumental variable estimators.

- The lagged $CO2_{t-1}$ variable is correlated with the error term ϵ_{t-1}. Moreover, the dataset is short in time (T=23). Hence, if a country's fixed effect is subject to a shock, this one would increase the correlation between the dependant variable and the error term, which increases the need of instrumental variables.

- The renewable energy consumption and the urban population are assumed to be endogenous, and the AB estimator allows us to take this specificity into account. Indeed, Inmaculada Martínez-Zarzoso highlighted in her paper an heterogeneous impact of urbanization on CO2, depending on the level of development. The renewable energy's endogeneity is also determined by this singularity (as it is possible to see with the Kuznet curve).

3 Results

In order to compare the AB estimator's results, we first estimate the models using OLS and fixed effects estimators. In the OLS regression, $CO2_{t-1}$ (co2L1) is positively correlated with ϵ_t, hence its coefficient is upward biased. Also, within transformation gives downward biased estimates (Nickell 1981). A good estimation should lies between these two models' results.

	(Model 1) OLS	(Model) F-E	(Model 2) OLS	(Model 2) F-E
co2L1	0.973***	0.562***	0.973***	0.564***
	(0.00509)	(0.0129)	(0.00510)	(0.0129)
gdp	0.00000730*	0.0000246***	0.00000748*	0.0000250***
	(0.00000293)	(0.00000562)	(0.00000293)	(0.00000579)
gdp2	-9.13e-11*	-2.87e-10***	-9.21e-11*	-2.90e-10***
	(3.90e-11)	(5.87e-11)	(3.89e-11)	(5.95e-11)
re	-0.00248***	-0.0107***	-0.00250***	-0.0108***
	(0.000506)	(0.00163)	(0.000505)	(0.00163)
up	-0.0000474	-0.00143***	-0.0000476	-0.00142***
	(0.0000291)	(0.000194)	(0.0000290)	(0.000194)
yr2009			-0.272***	-0.165**
			(0.0705)	(0.0615)
yr2010			0.109	0.109
			(0.0709)	(0.0620)
_cons	0.224***	2.328***	0.232***	2.323***
	(0.0337)	(0.0999)	(0.0341)	(0.0998)
N	3059	3059	3059	3059
adj. R^2	0.968	0.435	0.968	0.437

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Thus, we get a range for the estimator of $CO2_{t-1}$, that is $[0.973; 0.562]$ in the first model and $[0.973; 0.564]$ in the second model. The expected estimators should be between those values.

We don't impose a lag limit while doing xtabond2. The first two columns are one-step estimated, the last two are two-step estimated. We observe no significant difference in the estimated coefficients between one and two step estimators. But, indeed, the lag of CO2 coefficient fits in the range. In addition, coefficients are precisely estimated (significants with smalls standard errors). This method seems to perform well.

	(A-B 1step) Model 1	(A-B 1step) Model 2	(A-B 2step) Model 1	(A-B 2step) Model 2
L.co2	0.768***	0.774***	0.768***	0.774***
	(0.0668)	(0.0662)	(0.0668)	(0.0662)
gdp	0.0000300*	0.0000308*	0.0000299*	0.0000307*
	(0.0000126)	(0.0000129)	(0.0000128)	(0.0000128)
gdp2	-3.72e-10**	-3.79e-10**	-3.72e-10**	-3.78e-10**
	(1.17e-10)	(1.18e-10)	(1.17e-10)	(1.18e-10)
re	-0.00954*	-0.00997**	-0.00946*	-0.00990**
	(0.00383)	(0.00377)	(0.00377)	(0.00370)
up	-0.000645**	-0.000619**	-0.000646**	-0.000622**
	(0.000237)	(0.000236)	(0.000236)	(0.000237)
yr2009		-0.239***		-0.238***
		(0.0477)		(0.0479)
yr2010		0.141*		0.140*
		(0.0603)		(0.0601)
N	2894	2894	2894	2894
adj. R^2				

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.1 Diagnosis and lag sensitivity

The Arellano-Bond test confirms that there are no correlations of order two between error terms (A-B 1step Model1, same conclusions for others models).

Arellano-Bond test for AR(1) in first differences: z = -2.46 Pr > z = 0.014
Arellano-Bond test for AR(2) in first differences: z = -1.04 Pr > z = 0.300

Results are sensitive to the choice of number of lags. With T=23, we may loss in efficiency including all lags, so we could say that the third model (lags 1-8) appears to be a good compromise, even if re's coefficient is less precicely estimated.

	All lags	Lags 2-5	Lags 1-8
	co2	co2	co2
L.co2	0.768***	0.544***	0.670***
	(0.0668)	(0.0805)	(0.0700)
gdp	0.0000300*	0.0000349	0.0000281*
	(0.0000126)	(0.0000258)	(0.0000143)
gdp2	-3.72e-10**	-2.54e-10	-2.63e-10*
	(1.17e-10)	(1.84e-10)	(1.11e-10)
re	-0.00954*	-0.0122	-0.0104
	(0.00383)	(0.00723)	(0.00549)
up	-0.000645**	-0.00160***	-0.00103***
	(0.000237)	(0.000306)	(0.000269)
N	2894	2894	2894
adj. R^2			

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

However, while running a twostep regression of the models with lags restrictions (1-8), we get more precise estimators.

	A-B 2step Model 1	A-B 2step Mode2
L.co2	0.665***	0.673***
	(0.0786)	(0.0780)
gdp	0.0000284*	0.0000289*
	(0.0000119)	(0.0000117)
gdp2	-2.57e-10**	-2.60e-10**
	(8.81e-11)	(8.35e-11)
re	-0.0128**	-0.0135**
	(0.00479)	(0.00483)
up	-0.00113***	-0.00109***
	(0.000309)	(0.000308)
yr2009		-0.239***
		(0.0462)
yr2010		0.134**
		(0.0518)
N	2894	2894
adj. R^2		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Model 1 :

Arellano-Bond test for AR(1) in first differences: z = -2.69 Pr > z = 0.007

Arellano-Bond test for AR(2) in first differences: z = -1.08 Pr >z = 0.280

Model 2:

Arellano-Bond test for AR(1) in first differences: z = -2.65 Pr > z = 0.008

Arellano-Bond test for AR(2) in first differences: z = -1.23 Pr >z = 0.219

3.2 Results analysis and issues

We could confirm with no surprise that past CO2 levels influence future levels of CO2. We also catched the Kuznet Curve effect, with a significant and positive effect of GDP^2 on CO2 emissions, as well as the positive influence of the renewable energy consumption. Moreover, we catched the influence of the post-crisis period on CO2 emissions (due to the diminution of activity, underlined with the negative sign in front of the year 2009).

However, Inmaculada Martínez-Zarzoso find that "in uppermiddle income countries the elasticity, emission-urbanization, is negative", whereas it is a positive elasticity of 0.72 for less developed countries, and an elasticity superior to one for the most developed ones. Hence, to confirm this, one would need to separate the countries in order to classify them in homogeneous groups.
As an additionnal issue, we could say that the time dimension is not sufficiently long to capture all the effect of the urbanization and the use of the renewable energy, these two factors may vary slowly in time.

Bibliography Sahbi Farhani, *Renewable energy consumption, economic growth and CO2 emissions: Evidence from selected MENA countries*

C. Varotsos, M.-N. Assimakopoulos, and M. Efstathiou *Long-term memory effect in the atmospheric CO2 concentration at Mauna Loa*

Inmaculada Martínez-Zarzoso *The Impact of Urbanization on CO2 Emissions: Evidence from Developing Countries*

David Roodman, *How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata*

Annex1 Source data.worldbank.org (copy past)

CO2 Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.

GDP GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars.

Renewable energy consumption Renewable energy consumption is the share of renewables energy in total final energy consumption.

Urban population Urban population refers to people living in urban areas as defined by national statistical offices. It is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects.

Annex 2 Stata command :

```
xtset id,year
reg co2 co2L1 gdp gdp2 re up

reg co2 co2L1 gdp gdp2 re up yr2009 yr2010

xtreg co2 co2L1 gdp gdp2 re up, fe

xtreg co2 co2L1 gdp gdp2 re up yr2009 yr2010, fe

xtabond2 co2 L.co2 gdp gdp2 re up, gmmstyle(L.(co2) up re) iv(gdp gdp2) noleveleq robust

xtabond2 co2 L.co2 gdp gdp2 re up yr2009 yr2010, gmmstyle(L.(co2) up re) iv(gdp gdp2) noleveleq robust

xtabond2 co2 L.co2 gdp gdp2 re up, gmmstyle(L.(co2) up re) iv(gdp gdp2) noleveleq robust twostep

xtabond2 co2 L.co2 gdp gdp2 re up yr2009 yr2010, gmmstyle(L.(co2) up re) iv(gdp gdp2) noleveleq robust twostep

xtabond2 co2 L.co2 gdp gdp2 re up , gmmstyle(L.(co2) up re, laglimits(2 5)) iv(gdp gdp2) noleveleq robust

xtabond2 co2 L.co2 gdp gdp2 re up , gmmstyle(L.(co2) up re, laglimits(1 8)) iv(gdp gdp2) noleveleq robust

xtabond2 co2 L.co2 gdp gdp2 re up, gmmstyle(L.(co2) up re, laglimits(1 8)) iv(gdp gdp2) noleveleq robust twostep

xtabond2 co2 L.co2 gdp gdp2 re up yr2009 yr2010, gmmstyle(L.(co2) up re, laglimits(1 8)) iv(gdp gdp2) noleveleq robust twostep
```

Annex 3 Angola Albania Argentina Armenia Australia Austria Azerbaijan Burundi Belgium Benin Burkina Faso Bangladesh Bulgaria Bosnia and Herzegovina Belarus Belize Bolivia Brazil Bhutan Canada Switzerland Chile China Cote d'Ivoire Cameroon Congo, Rep. Colombia Comoros Cabo Verde Costa Rica Cuba Czech Republic Germany Dominica Denmark Dominican Republic Algeria Ecuador Egypt, Arab Rep. Eritrea Spain Estonia Ethiopia Finland Fiji France Gabon United Kingdom Georgia Ghana Honduras Croatia Haiti Hungary Indonesia Ireland Iran, Islamic Rep. Iraq Iceland Italy Jamaica Jordan Japan Kazakhstan Kenya Kyrgyz Republic Korea, Rep. Lao PDR Lebanon Liberia Sri Lanka Lithuania Luxembourg Latvia Morocco Moldova Madagascar Mexico Macedonia, FYR Mali Myanmar Mozambique Mauritania Mauritius Malawi Malaysia Namibia New Caledonia Niger Nigeria Nicaragua Netherlands Norway Nepal New Zealand Pakistan Panama Peru Philippines Palau Papua New Guinea Poland Portugal Paraguay Pacific island small states French Polynesia Romania Russian Federation Rwanda South Asia Sudan Senegal Singapore El Salvador Sub-Saharan Africa Small states Sao Tome and Principe Suriname Slovak Republic Slovenia Sweden Swaziland Syrian Arab Republic Togo Thailand Tajikistan Turkmenistan Trinidad and Tobago Tunisia Turkey Tanzania Uganda Ukraine Uruguay United States Uzbekistan St. Vincent and the Grenadines Venezuela, RB Vietnam Samoa South Africa Congo, Dem. Rep. Zambia