European Commission

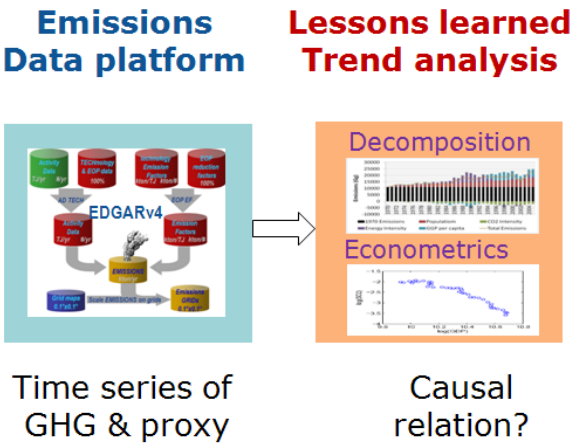
JRC TECHNICAL REPORTS

Analysis of Greenhouse Gas Emission Trends and Drivers

A Survey of Techniques for Emission Decomposition and Econometric Trend analysis

Greet Janssens-Maenhout (JRC-IES-H02-EDGAR)
Paolo Paruolo (University Insubria/ JRC-IPSC-G03)
Simone Martelli (JRC-IES-H02-EDGAR)

Acknowledgement to:
Vivien Tran-Thien (ex-CLIMA)
Velina Pendolovska (CLIMA)



2013

Report EUR 25814 EN

European Commission
Joint Research Centre
Institute for Environment and Sustainability (H02 – EDGAR Emissions team)
Institute for Protection and Security of the Citizen (G03 – CRIE Econometrics team)

Contact information
Greet Janssens-Maenhout
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 290, 21027 Ispra (VA), Italy
E-mail: greet.maenhout@jrc.ec.europa.eu
Tel.: +39 0332 78 5831
Fax: +39 0332 78 5704

http://edgar.jrc.ec.europa.eu/aetad/index.php
http://www.jrc.ec.europa.eu/

This publication is a Reference Report by the Joint Research Centre of the European Commission.

Legal Notice
Neither the European Commission nor any person acting on behalf of the Commission
is responsible for the use which might be made of this publication.

Europe Direct is a service to help you find answers to your questions about the European Union
Freephone number (*): 00 800 6 7 8 9 10 11
(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu/.

JRC 78707

EUR 25814 EN

ISBN 978-92-79-28602-5 (pdf)
ISBN 978-92-79-28601-8 (cd-rom/dvd)

ISSN 1831-9424 (online)

doi: 10.2788/84075 (online)

Luxembourg: Publications Office of the European Union, 2013

© European Union, 2013

Reproduction is authorised provided the source is acknowledged.

ANALYSIS OF GREENHOUSE GAS EMISSION TRENDS AND DRIVERS (AETAD)

REPORT 1: Literature survey and project outlook

Deliverable 1 under Administrative Arrangement n. 071201/2011/605783/CLIMA.A3

July 27, 2012

History of the document

Date	*Authors/contributors*	*Filename*	*Version*
08/05/12	G.Janssens-Maenhout, P. Paruolo	AETAD_080512_report1	First draft submitted for comments
27/05/12	V. Tran-Thien		Revision by DG CLIMA
16/07/12	G.Janssens-Maenhout, P. Paruolo	AETAD_report1_final	Revised report submitted for approval
23/07/12	V. Tran-Thien, V. Pendolovska		Second revision by DG CLIMA
27/07/12	G.Janssens-Maenhout, P. Paruolo	AETAD_report1	Final report

European Commission, Joint Research Centre
Institute for Environment and Sustainability, Air and Climate Unit, EDGAR team
Via Fermi 2749, TP 290
I-21020 Ispra (VA) Italy

Contents

Abstract

This first report of the AA on "Analysis of greenhouse gas emission trends and drivers" for DG CLIMA contains the literature review on emission trends and drivers, and describes two common methods of trend analysis:

1. decomposition methods for defining decomposed indicators,
2. econometric techniques for finding cause-and-effects links and quantitative relationships between emissions and drivers.

Given the dynamic importance when assessing the impact of drivers on emissions in near-term scenarios, the methodology proposed in this report is to use econometric techniques as in 2., possibly in combination with decomposition methods in 1., using time-series data.

Concerning data, an appropriate level of disaggregation within sectors and even sub-sectors is required for the analysis. The potential for future policy making is highest in the non-ETS sectors, such as road transport and households sectors.

The application of the methodology demands that appropriate time-series data is available for the EU-27 Member states. Such data needs to be carefully assembled from different sources. The availability of data for EU countries appears significantly lower than for the United States, where data of good quality is abundant.

Given the scope of the present project, a selection of a pilot case study appears mandatory. The pilot case study will be used to illustrate the potential benefits of the proposed methodology, which will involve econometric methods (as in 2.) possibly combined with decomposition methods (as in 1.).

Chapter 1
Introduction

1.1 Background

The European Council committed in March 2007 to the 20-20-20 targets for 2020 of at least 20% greenhouse gas (GHG) emission reduction below 1990 levels, 20% of energy from renewables and 20% increase in energy efficiency. Thereto a set of legislative acts were established under the Climate and Energy Package aiming at meeting these targets, but the provisions of the Climate and Energy Package do not explicitly aim at reaching the +20% energy efficiency target. This is accompanied by strengthening of the ETS, the Effort Sharing Decision for non-ETS sectors, the renewable energy share in the Member States and the promotion of Carbon Capture and Sequestration.

Moreover, the IPCC (2007) analysed the consequences of an increase of world average temperature by several degrees (global warming). It introduced the 2°C target as maximum average temperature increase above pre-industrial levels and it described the corresponding need of mitigating the current trend of GHG emissions. The agreements reached at the UNFCCC climate negotiations in Copenhagen in 2009 adopted internationally this target of limiting average global temperature increase to 2 °C above pre-industrial levels. Also furtheron, the negotiations in Cancún in 2010 and Durban in 2011 indicated that all countries should take urgent action to reduce global greenhouse gas emissions, in order to limit the increase in global average temperature to this 2 °C target.
Since 2000, an estimated total of 420 billion tonnes CO2 was cumulatively emitted due to human activities (including deforestation), as reported by Olivier et al. (2012). Meinshausen (2009) suggested that not exceeding the 2 °C target is possible if cumulative emissions in the 2000–2050 period do not exceed 1,000 to 1,500 billion tonnes CO2. If the current global increase in CO2 emissions continues, cumulative emissions will surpass this total within the next two decades, well before the year 2050.

Therefore, over the longer term (2050) the European Union has produced the 2050 Roadmap, as communicated by the European Commission, that addresses the remaining GHG emission problems. An option remains a decarbonisation of the European (urban) society which includes e.g. reducing the transport and household sector emissions; a reduction of 80-95% below 2005 levels is proposed.

Developing and implementing climate change policy requires a sound analysis of past GHG emissions. Such an analysis can help to detect a risk of deviation from emissions reduction targets, to identify the emission sources to target in priority, to evaluate the effect of mitigation policies and to improve emissions projections.

Although there can be general agreement regarding the identification of the main drivers of the emissions (such as economic activity, fuel-efficient technologies, climatic conditions…), the quantitative relationships linking them to the emissions have not been systematically and comprehensively investigated. Ultimately these drivers are a key for effective emission reduction policies.

Even though many industrial emissions sources are largely capped with corresponding carbon tax under the Emissions Trading System (ETS), others, in particular those not under the ETS, are less understood, the effect of drivers is not precisely quantified, and they are much less controlled. As such, the non-ETS sectors bear still a large potential for emission reduction.

1.2 Objective of the project

The objective of this project is the development of a method to systematically and quantitatively analyse the influence of the main emission drivers on past emissions trends. As opposed to studies based only on a decomposition analysis, and in order to offer complementary insights, this project discusses the advantages and limitations of an econometric approach.

Overall, the project will help:
(i) to establish a methodological approach for analyzing emission trends,
(ii) to understand how to identify the drivers of GHG emission trends,
(iii) to develop effective mitigation instruments at sector-specific level.

1.3 Selected methodology

The European Commission is interested in understanding the links between changes in drivers and emissions. Hence it is of interest to know the potential some drivers bear for reducing emissions. This project has a more limited scope:

1) the first part of the project consists of a literature review; this provides an overview of the existing decomposition and econometric analyses that are currently applied to aggregate and disaggregate data on GHG and potential drivers for different world countries or regions;

2) the second part of the project describes a systematic econometrics approach, potentially valid for all world-countries, that can be used to test the relationship between potential drivers and emissions. This approach is based on cause-effect econometric models that allow

A) to ascertain if a certain driver has proved historically to be a significant cause of emissions,
B) (in case of significance) to measure quantitatively the impact of the driver on emissions,
C) to provide measures of the uncertainty associated with this relation. The uncertainty analysis allows to derive uncertainty bands for emissions predictions based on given scenarios on the future development of the drivers.

The feasibility of this cause-effect econometric model will be demonstrated on some pilot case study, selected from one non-ETS sector, such as road transport (TS) or household (HS).

The selection of the econometric techniques is based on the most appropriate practices in the literature, as well as on a previous exploratory research of the project team. Recently, Paruolo et al. (2011) carried out a statistical analysis of correlation between emissions and income trends on global scale on EDGAR data. This analysis showed some of the benefits of the application of econometric methods.

The very different trends for different sectors (e.g. transport, power generation, industry and buildings) suggest a disaggregated sector-by-sector approach. A country-specific approach seems also appropriate, given different institutional and technological differences among member states, also within a given sector. While a complete sector-and-country specific analysis is beyond the scope

of the present analysis, the project will illustrate the proposed methodology on a pilot case study, both at a country level and at EU 27 level. The analysis can be extended to other sectors once lessons from the pilot case study can be drawn.

The proposed analysis will apply an econometric approach, possibly combined with information based on decomposition analysis. The change of the log of the Kaya identity (or of similar identities[1]) provides a framework for the different drivers contributing to emission trends at sector-specific level. This information can be coupled with an econometric cause-effect model on each of the variables that appear in the Kaya identity. The explanatory variables (causes) in these models may include potential policy tools, as well as other potential causes, such as fuel prices for the TS. This approach allows subsequently to aggregate the predictions of each variable in the Kaya identity into an aggregate prediction on emissions.

1.4 Structure of this report

The rest of this report is organized as follows. Chapter 2 gives a literature overview of the decomposition analyses. Chapter 3 describes relevant econometric studies. In the final Chapter 4, we summarize the literature review and present the outlook on what lessons can be learned from the previous two chapters.

We find clear arguments in favor of the use of the econometric methods, also because of the dynamic character of the impact assessment of drivers on emissions. We also argue that it is possible to exploit decomposed indicators as potential drivers. These are complemented with other potential drivers suggested in the econometrics studies. The proposed methodology for this project is also discussed; we propose to test statistically all potential drivers with econometric time-series techniques, and subsequently to derive sector-specific, econometric relationships between emissions and drivers.

[1] See Chapter 2 on Decomposition analysis.

Chapter 2

Decomposing emission inventories

2.1 Common background knowledge on emission indicators of IPCC

Some of the major driving forces of past and future anthropogenic greenhouse gas (GHG) emissions, which include demographics, economics, resources, technology, and (non-climate) policies, were already reviewed by the IPCC, see Nakićenović, N., and R. Swart (2000). Major driving forces can be structured by considering the links from demography and the economy to resource use and emissions, leading to an identity of

*Impact = Population * Affluence * Technology level*

where *Affluence* is usually measured as income *per capita* and *Technology level* is usually measured as emissions per unit of income. This identity has been extensively applied in analyses of energy-related CO2 (e.g., Kaya (1990), Ogawa (1991), Nakicenovic et al. (1993), O'Neill et al. (2000)) and is often referred to as the *Kaya identity*:

CO2 = *Population* * (GDP/*Population*) * (*Energy*/GDP) * (CO2 /*Energy*)

or

CO2 = *pop * income * EI * CI,*

where the second equation introduces acronyms for the effects in the first equation. A property of the multiplicative Kaya identity is that component growth rates are additive. In fact, taking logarithms and subtracting values for time t and for time t-1, one obtains

$$\Delta \log(CO_{2t}) = \Delta \log(CI_t) + \Delta \log(EI_t) + \Delta \log(income_t) + \Delta \log(pop_t)$$

where $\Delta a_t = a_t - a_{t-1}$ indicates the first difference (growth) of a_t. We refer to this identity as *the Kaya identity in growth rates*.

For instance, global energy-related CO2 emissions since the middle of the 19th century are estimated to have increased by approximately 1.7% per year. This growth rate can be decomposed roughly into a 3% growth in gross world product (the sum of a 1% growth in population and a 2% growth in per capita income) minus a 1% per year decline in the energy intensity of world GDP and a decline in the carbon intensity of primary energy of 0.3% per year (Nakicenovic et al., 1993).

While the Kaya identity above can be used to organize discussion of the primary driving forces of CO2 emissions and, by extension, emissions of other GHGs, there are important caveats. Most important, the four factors on the right-hand side of the Kaya equation should be considered neither as fundamental driving forces in themselves, nor as generally independent from each other. Global analysis[2] is often not instructive and even misleading, because of the great heterogeneity among populations and level of industrialization with respect to GHG emissions.

[2] Global analysis here stands for the joint analysis of all world countries.

Long-run per capita economic growth and structural change are closely linked with advances in knowledge and technological development and drive capital and labor productivity. Most innovative efforts in the past two centuries were devoted to improving labor productivity and the human ability to harness resources for economic purposes. While material and energy efficiency improved slowly, economic growth was faster and thus the aggregate use of resources increased. Progress in the development of clean technology is expected as incomes rise, a relationship often referred to as the "Environmental Kuznets Curve" (EKC). Even though this process has possibly been documented for sulfur emissions, it is not proven for GHG, cfr. a.o. Viguier (1999) or Baek et al (2009).

Box1: What is an emission decomposition analysis?

The emissions decomposition analysis consists of an *accounting identity* (similar to the ones used in *National Accounts*, see World Bank (2008) for a definition). The identity decomposes emissions into component indicators, in order to describe the driving forces of emissions in a given inventory. The resulting accounting identity can be multiplicative or additive.

Example: A well-known identity is the identity of Kaya (1993), which decomposes emissions multiplicatively into population, income, energy intensity and carbon intensity:

$$CO2 = Population * (GDP/Population) * (Energy/GDP) * (CO2 /Energy)$$

Pros: being an identity, there is no residual error associated with decomposition analysis.

Cons: the identity gives a breakdown of the (annual) emission inventory, without explicitly analyzing the causes and the dynamics of the emission trends.

2.2 Available decomposition analyses

At the G8 Summit in Summer of 2005 the Gleneagles Plan of Action for clean energy and sustainable development was launched, for which the IEA developed indicators to provide data and analysis on energy use, efficiency developments and policy pointers. The IEA/OECD (2007a) analysed the energy use and CO2 emissions of 14 OECD member countries from 1990-2004 using the Laspeyres index method and energy indicators. It confirmed the conclusions of previous IEA (2004) study that the changes caused by the oil price shocks in the 1970s and the resulting energy policies did considerably more to control growth in energy demand and reduce CO2 emissions than the energy efficiency and climate policies implemented since the 1990s. The energy indicators are specified for each energy-related activity (and industrial subsector) by IEA (2007b). The applied IEA method used the Laspeyres Index method, i.e. the index number formula as developed by Laspeyres, originally developed for price indexation. The Laspeyres indices calculate a weighted average of a compound variable at time t with its share at time 0. The application of this Laspeyres method was analysed by Ang and Liu (2007) and some readjustments were suggested to account for the large changes in structure and energy intensity in developing countries.

At European level, evidence-based policy making as well as proper monitoring of the efforts to reduce emissions require more in-depth analysis of the drivers of emissions. The European Environment Agency EEA (2010) presented a decomposition of the 1990, 2000 and 2008 emissions

per Member State. In the EEA study, all sectors are analysed separately; within each sector, additive and multiplicative decompositions were applied for the sake of simplicity. Even though also non-energy related sectors were treated, the report did not follow a particular scientifically-recommended decomposition method similar to the IEA method.

For the energy-related sectors, a purely multiplicative method could have been applied, using the Energy efficiency indicators[3] for Europe from ODYSSEE-MURE (2010) of and taken up in Table 5 and 6 of Annex 2, which is usually done for the TERM (Transport & Environment Reporting Mechanism) report of EEA (2009). Under the Intelligent Energy Europe Programme, the ODYSSEE-MURE project (2009) gathered data and indicators to monitor energy efficiency (ODYSSEE – online database for Europe's Energy Efficiency) and policy measures (MURE- Policy Measures for energy efficiency) for EU-27, Norway and Croatia. The project aimed at understanding the energy trends and at measuring the contribution of innovative energy efficiency and renewables. The outcome is also used by EuroSTAT, but the data remains limited with relative short time-series and a limited cross sectional (country-specific) data.

To enhance the cross-sectional data, the decomposition methods used by some other OECD countries can be considered useful: the Australian Bureau of Agricultural and Resource Economics (ABARE) uses the residual free Lon-Mean Divisia Index decomposition method for energy intensity trend analyses; Natural Resources Canada uses the same method to track its energy trends; the Japan Institute of Energy Economics provides yearly data on energy and economics, highlighting key indicators; the US Department of Energy uses the logarithmic mean Divisia decomposition methodology with index aggregation across the hierarchy of indicators; the Energy Efficiency and Conservation Authority of New Zealand reviews every 5 year efficiency progress with a Divisia decomposition to separate the effects of activity, structure, quality, weather and technical efficiency.

2.3 Perfect decomposition techniques in energy and environmental analysis

Perfect index decomposition method techniques do not give rise to residuals. Not all index decomposition method techniques are perfect, see Ang and Liu (2007). As an example of perfect decomposition method, assume that total energy consumption in year 1 and year 2 are, respectively, given by E1 and E2. In the IEA model, the change in energy consumption from year 1 to year 2, given by (E2/E1), is decomposed into component indices associated with the activity effect, the structure effect and the intensity effect. The residual term is then given by dividing (E2/E1) by the product of the indices of these three effects. For perfect decomposition, the residual term is equal to unity. The larger the deviation of the residual term from unity, the greater is the change in total energy consumption that is unexplained.
The perfect index decomposition method techniques promoted by Ang and Liu (2007) are:

- Fisher: modified Fisher ideal index method,
- AMDI: arithmetic mean Divisia index method,
- LMDI-1: log mean Divisia index method 1
- LMDI-2: log mean Divisia index method 2.

Ang (1999) studied decomposed indicators for 10 world regions with temporal and cross-sectional comparisons. The variation attributed to factors as production mix of economy or industry, fuel mix and energy efficiency is what he defines as decomposition analysis with carbon intensity and carbon factor as commonly decomposed indicators. Special boundary cases of non-CO2 emitting sectors are treated by Ang (2002).

[3] For the more explicit formulation of these ODYSSEE indicators, refer to Tables 5 and 6 of Annex 2.

Other decomposition techniques, based on the Shapley value[4] of game theory, can be used to study CO2 emissions in four OECD countries with a decomposition without "unexplained" residuals, as demonstrated by Albrecht (2002) and Sun (1998). The Shapley decomposition refines the Laspeyres index method in order to have very small residuals.

The choice of one of the index methods depends strongly on the availability of the underlying activity data required differently for the different index methods (Fisher, AMDI, LMDI-1 ad LMDI-2).

2.4 An application of decomposition techniques

The EDGARv4.2 database is compiled bottom-up with a technology-based methodology using international statistics to the extent possible for the activity data and internationally accepted emission factors (from IPCC 2006 guidelines or EMEP/EEA 2009 guidebook). Therefore EDGARv4.2 can be considered to be constructed with decomposed factors (Activity data * technology share * emission factor * end-of-pipe reduction) and is as such suited for decomposition, by e.g. the refined Laspeyres indices method, a technique recommended by Ang et al. (2009).

In order to illustrate the use of decomposition techniques in this project, consider the Kaya identity in growth rates. Fig. 1 reports the results of such analysis with the EDGARv4.2 data for CO2, data from IMF (2011) for income, and data from UNDP (2011) for population. The figure shows the large variation over time in growth of CO2 emissions and income for three large economic regions: USA, EU-27 and China. The variation of population dynamics appears comparably much smaller and much smoother; this suggests that population dynamics are more predictable on the basis of past data but that it has a limited impact on the prediction of CO2 growth.

Comparing world regions, the variation of CO2 in China has a higher average growth, reflecting the steep increase in CO2 emissions, than the USA and EU-27, which show a similar behaviour, both for the growth of CO2 and of income. Remark that the recession of 2008-2009 is visible for all the three regions, where the strongest dip over the 1990- 2008 period in the growth of income and CO2 can be observed for the USA and EU-27.

[4] The Shapley value is a solution concept in cooperative game theory, assuming a unique distribution amongst the commons of the problem, with the following setup of a coalition of commons cooperating and obtaining a certain overall gain from that cooperation. The Shapley value provides an answer to how important each common is to the overall cooperation.

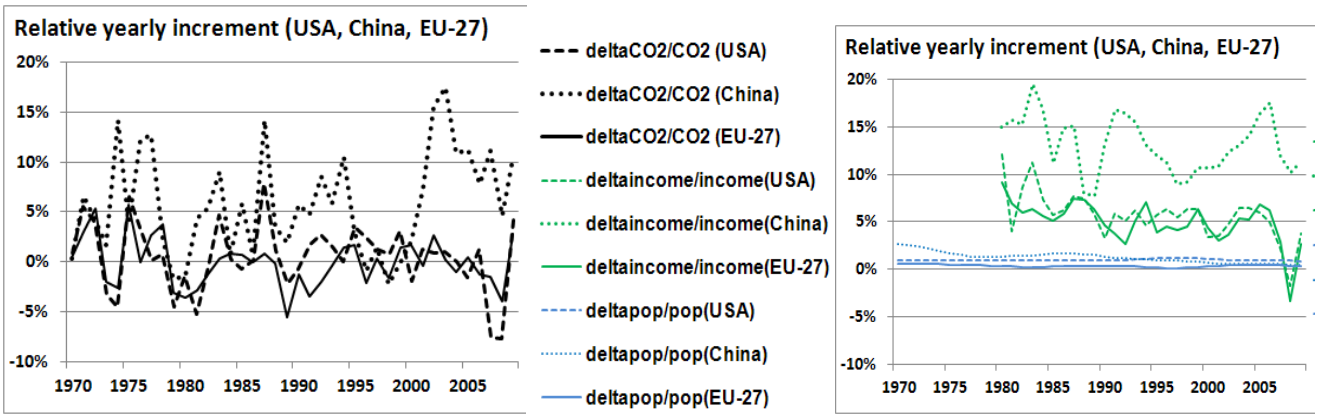


Fig. 1: Delta log of CO2 emissions (a) and of some decomposed indicators (b) – income and population for three world players: USA, China and EU-27.

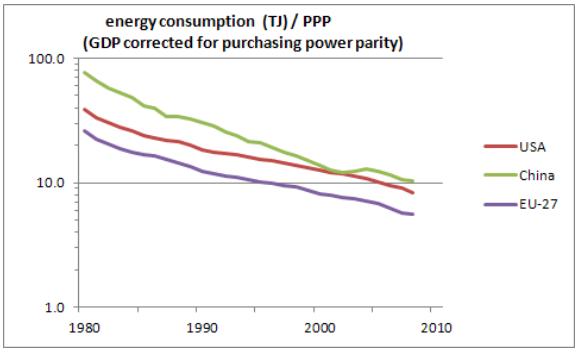


Fig. 2a: Time-series energy consumption over the purchasing power parity.

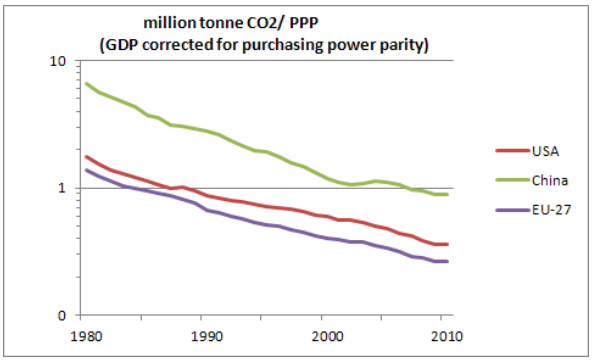


Fig. 2b: Time-series evolution of the CO2 emissions over the purchasing power parity. This ratio shows in the log diagram a steady decrease and can be described as an exponential decay with EU-27 as the smallest emitter, followed closely by USA. China is approaching the situation of EU-27 in the early eighties.

This observation suggests to concentrate the econometric modeling analysis on CI and EI and of their determinants. Even though the project focuses primarily on the understanding of past emission trends, it bears the possibility to predict near-term future changes. The CI and EI can then be coupled with income and population predictions through the Kaya identity in order to obtain predictions on the growth of CO2 emissions.

The energy-related sectors are the most important determinants for the dynamics of emissions and of income, which is demonstrated in fig. 2a. This is also reflected in the trend of CI*EI= CO2/income observed in Fig. 2b, which is highest for China (characterized by high energy consumption per GDP) and lowest for EU-27.

The use of the Kaya identity illustrates the benefit of using decompositions: they help to understand the driving forces of emissions by decomposing the emissions into contributing factors that require different treatments to analyse their interconnections. In the example of the Kaya identity:

- population dynamics can be predicted with simple extrapolation methods, given its small variability;
- scenario on the evolution of predictions on income are often available, and do not require ad-hoc modeling;
- the predictions of EI and CI are instead the key areas of analysis, and they require appropriate ad-hoc econometric modeling, see the following sections. These econometric models allow to evaluate the impact of different policy measures, as well as the effects of likely scenario on the evolution of other drivers that are not under control of the policy-maker.

For simplicity, we have just considered the Kaya identity, which is here defined at aggregate level for all energy-related sectors together. Decompositions can be also considered at disaggregate level, as illustrated in Section 3.4 below. We now turn to the description of the econometric analysis.

Chapter 3

Modelling causes and effects

3.1 Common knowledge on emission trends and future scenarios by IPCC

For the Fourth Assessment Report (AR4) of the IPCC, Meehl et al (2007) selected three climate scenarios, or so-called storylines for the 21st century from the IPCC Special Report on Emissions Scenarios of Nakicenovic and Swart (2000). The latter Special Report describes the following as major driving forces of past and future anthropogenic GHG emissions: demographics, economics, resources, technology, and (non-climate) policies.

It also refers to the Enviromental Kuznets Curve (EKC) as a process that seems well established for traditional pollutants, such as particulates and sulfur. The EKC is a non monotonic relation between emissions as a function of income, which is assume to have positive slope for low values of income and negative one for large values of income. Nakicenovic and Swart (2000) mention that the EKC might apply to GHG emissions but are aware that Stern (2004) and Wagner (2008) amongst others argued that the Kuznets curve is not corroborated by robust statistical evidence on GHG emissions. Notwithstanding these criticism, the main drivers in the overview table for all storyline scenarios are primarily linked to *per capita* income.

Box2: What is an econometric equation?

An *econometric equation* (*or behavioral relation*) usually takes the form

$$y_t = \alpha + \beta x_{1t} + \gamma x_{2t} + \varepsilon_t$$

where $\alpha + \beta x_{1t} + \gamma x_{2t}$ is assumed to represent the average response of y_t given x_{1t} and x_{2t}. The error term ε_t is random, and it is assumed to be zero on average, so as to make $\alpha + \beta x_{1t} + \gamma x_{2t}$ equal to the average response of y_t given x_{1t} and x_{2t}.

Example In the analysis of emissions trends, one may be interested in explaining the amount of gasoline demanded at time t, q_t say. In this case y_t may be taken to be $\Delta \log q_t$, and x_{1t} equal to $\Delta \log income_t$, and x_{2t} equal to $\Delta \log p_t$, where p_t is the price of gasoline. Note that one may take x_{2t} equal to $\Delta \log p_{t-1}$, making the equation a dynamic one.

If $\gamma \neq 0$ one can say that x_{2t} is a driver of y_t. When $x_{2t} = \Delta \log p_{t-1}$, one says that prices p_t Granger-cause gasoline consumption q_t.

Pros: one can test if x is a driver of y, also in a dynamic model. Granger-causality can be addressed in a statistically robust manner and near-term prediction models can be derived.

Cons: as the dynamic equation is not an identity, it includes an error which represents the uncertainty in the y-x relation.

Reference For a detailed discussion on the interpretation of the coefficients in econometric equations, in system of equations and in cointegrated systems see Johansen (2005).

3.2 Analysing causality

A systematic econometric approach can be used to test the relationship between drivers (X) and the derived indicators (Y) in the decomposition analysis. In the example of the Kaya identity above, an example of derived indicators is EI and an example of driver is fuel price. The derived indicators from decomposition analyses can be used as potential drivers. This list can be complemented with other drivers identified in econometric studies.

This approach is based on cause-effect econometric models that allows

A. to ascertain if a certain driver X has proved historically to be a significant cause of Y,
B. (in case of a significant influence of X on Y) to measure quantitatively the impact of the driver X on Y,
C. to quantitatively measure the uncertainty associated with the relation between variations in X and variations in Y.

Note that part C) is of interest both in case when driver X is a policy variable and in case when X is not a policy variable but a likely scenario on its future development can be foreseen. In the following we refer to these steps as the ABC steps. The feasibility of this cause-effect econometric model will be demonstrated on a pilot case study.

The specification of the econometric techniques required in this part of the project rests on a large scientific literature on the investigation of causes and effects. In this context *statistical significance* plays a key role: if X is not found to have statistical significance in the explanation of Y, then there is no sufficient scientific evidence of a cause-and-effect nexus between X and Y. The scientific literature on causality has deep roots, and we refer to Pearl (2009) for a review.

These techniques have been introduced and fruitfully applied in econometrics to the analysis of cross-section data, or also of panel data, see Wooldridge (2011) for a general reference.[5] An example for which cross-sectional data would be relevant for the present project in the case of HS; usually information on substitution effects and price elasticities are estimated on data collected through interviews of individual households, in which a certain number of micro-variables are collected. An example of these variables includes: quantity of energy type consumed by household, price of energy type, income of the household, number of members of the family, type and size of house etc. Section 3.2 below illustrates state-of-the art econometric techniques used in this area.

The present project is targeted at decomposition and modeling of time-series data. For time-series, the discovery of (i.e. inference on) a cause-and-effect nexus is usually associated with the notion of "Wiener-Granger causality" or more simply "Granger causality", see e.g. Geweke (1984).The main idea behind Granger-causality is that X is said to Granger-cause Y if the prediction of Y(t+h) at time t for some positive horizon h, on the basis of information on Y(t-m) with m=0,1,2,… can be improved by knowledge on X(t-m), with m=0,1,2,… Here, X(t) indicates X at time t, and similarly, Y(t) indicate Y at time t. In words, X Granger-causes Y if it can help predict Y.

[5] A useful classification of data types is the following: **cross-section** data, **panel data**, **time-series** data. A data set is defined to be a **cross-section** if it consists of data on different units (like households or countries) for the same time period. A **time-series** data set contains instead data on one or more units for a relatively large number of time periods (40 time periods, say). Finally, it is defined as a **panel data** set if it consists of data on different units for a small number of time periods. This classification is not clear-cut, and it is mainly used to refer to the statistical models that can be applied to the different types of data.

This concept is directly applicable to the analysis of the nexus between the time-series of X and Y in point A) above. It should be emphasized that results of a Granger-causality test can be either significant or not. In the former case, the test supports the use of X as a predictor ("driver") for Y; in the latter, it suggests to discard the X as a predictor for Y. This lack of significant effects can be also due to insufficient variation or information present in the data. Because the data is generated by a so-called natural experiment (as opposed to a controlled experiment), researchers can just perform significance tests on the available, given data and act accordingly; however they cannot generate more data to resolve the indecision implied by the insignificance of the test, see e.g. Haavelmo (1944). This feature is common to all non-experimental sciences.

Time-series data are often trending over time. As such they can generate spurious results, if standard statistical tests are applied to them. This problem has a long history, and it is known as the *spurious regression problem*, see Yule (1926). In order to investigate the nexus between trending time-series, Engle and Granger (1986) have introduced the concept of *cointegration* which is dual to the idea of *common trends* between time-series. This concept has proved very useful in modeling relations involving non-stationary time-series. Cointegration delivers a predictive equation of growth rates, such as $\Delta\log(CI_t)$ or $\Delta\log(EI_t)$ in the Kaya identity in growth rates above; this predictive equation is called "Error correction model" (ECM), see Hendry (1995). Nowadays, Granger-causality, cointegration and error correction are all coupled together in the specification analysis of relationships involving trending time-series, see e.g. Engle and White (1999). These techniques are current state-of-the-art in the analysis of emission trends.

An example of the use of cointegration techniques is given by the analysis of the nexus between income and emissions developed by some of the team members, see Paruolo et al. (2011). They use data from EDGARv4.1 dataset for emissions of CO2 and SO2 and the Penn World Table 7.0 data for GDP. In line with the standard exposition of the EKC theory, this paper assumes that the EKC relation can be approximated by a linear relation between emissions and income over periods comprising a few decades. The EKC predicts that the slope of this relation is positive for developing economies and negative for fully developed ones.

The paper analyses each countries separately and looks for cointegration, as a sign of presence of the EKC. Only for a minority of countries such a relation is found to exist. For these countries, one could assume that a unique, homogeneous EKC curve exists. The paper hence tests for the presence of such an homogeneous EKC curve that links countries' income and the slope of the linear approximation to the EKC. The paper finds that this relation is insignificant.

We note here that the assumption of a homogeneous EKC curve is a standard assumption in studies of the EKC based on cross-section regressions and on panel data. Moreover, it is a standard tenet of policies inspired by the EKC, which prescribe to help developing countries reach the "turning point", i.e. the level of income above which the slope becomes negative. Many of these studies (try to) estimate this (single) turning point, by pooling data on all countries; hence they assume implicitly that a single EKC curve exists for all countries. Remark that in Paruolo et al. (2011) the homogeneity assumption of the EKC is only made on the last stage of the analysis for the countries which display cointegration.

The results in Paruolo et al. (2011) are based on the cointegrated model, and they challenge the main implications of the EKC hypothesis, i.e. that with increasing income, environmental degradation diminishes. As a result, policies imposing control measures on industry and citizen are needed even more in order to prevent or mitigate environmental degradation. This example also illustrates how these methodologies have direct implication for policy.

In the following subsections we illustrate prototype econometric models and techniques for cross sectional and time-series data, both for the HS and the TS sectors. Under the HS emissions we consider the direct emissions of heating of buildings.

3.3 ABC for cross-sectional data – Household sector

In this section we illustrate a reference model for the analysis of cross sectional and/or panel data. In order to make the exposition concrete in this section, we consider the example of the household sector (HS), testing for the effect of energy prices on Household energy demand.

Here we follow the specification in Berkhout et al. (2004) for a small demand system for energy based on cross sectional data. This example exemplifies steps A), B) and C) from the previous section in the present context. The role played by the type of available data is emphasized.

The purpose of Berkhout was to perform an ex-post *impact assessment* of the effect on an energy tax introduced in the Netherlands in 1996. The paper concluded that the average demand reduction was 8% for electricity and 4.4% for gas, and compares this with counterfactual reductions that could have achieved through alternative policies.

Data
The study was performed on a number of households (in the range of 1000 to 1500) from the NL, which was monitored for the years 1994 to 1999. This type of data is called a *panel*. The same data for a single year is called a *cross-section*. The time dimension in this study was mainly used to take first differences across successive years, and we omit to indicate it explicitly in what follows.

The available data can be grouped in the following categories

Energy demands: Let x_{ih} indicate the quantity (demand) for energy type i by household h, where $i = e, g$ where e stands for electricity and g for gas.

Income: Let y_h be household income.

Prices: Let p_i indicate price of energy type i, p^{en} the price of energy (a geometric weighted average of p_e and p_g), p^{ne} the price of non-energy goods, P the general price level (a geometric weighted average of p^{en} and p^{ne}).

Household characteristics: income and behavioral variables: income, family size, showering and cooking behavior etc. These variables are collected in a vector called D_h

House characteristics: house size and type, ownership of durable goods and electrical appliances such as dishwasher, laundry-drier etc. These variables are collected in a vector called B_h. They also included average winter temperature, which was deemed important for demand for gas.

From the above data, define the following derived variables:

$\tilde{p}_{ih} := p_i/y_h$, the price of energy type i divided by household income;

$s_{ih} := x_{ih}/y_h$, the demand "share" of energy i for household h.

Specification
The specification for the small demand systems for energy by households is given in eq. (7) of the paper, and it can be rewritten here as follows:

$$s_{eh} = \alpha_{1e}\tilde{p}_{eh} + \alpha_{2e}D_h + \alpha_{3e}\ln\left(\frac{p^{en}}{p^{ne}}\right) + \alpha_{4e}\ln\left(\frac{y_h}{P}\right) + \alpha_{5e}B_h + \varepsilon_{eh}$$

$$s_{gh} = \alpha_{1g}\tilde{p}_{gh} + \alpha_{2g}D_h + \alpha_{3g}\ln\left(\frac{p^{en}}{p^{ne}}\right) + \alpha_{4g}\ln\left(\frac{y_h}{P}\right) + \alpha_{5g}B_h + \varepsilon_{gh}$$

The dependent variables of the system are $Y_h := (s_{eh}, s_{gh})$. The explanatory variables are $X_h := (\tilde{p}_{eh}, \tilde{p}_{gh}, D_h', \ln\left(\frac{p^{en}}{p^{ne}}\right), \ln(\frac{y_h}{P}), B_h')'$. Here ε_{eh} and ε_{gh} are stochastic errors, that are typically assumed to be uncorrelated with the explanatory variables in X_h .

α_{ji} are unknown parameters to be estimated on the available data.

The main features of the system that make it a demand system are:

- it relates energy demand (shares) to prices of the same energy source as well as of alternative energy and of alternative goods; this allows e.g. to compute the effect of a change in relative prices on the two energy demands;
- it relates energy demand (shares) to income or total consumption; this allows e.g. to compute the effect of a change in income on energy demand.

The system also relates energy demand to households and house characteristics; this allows e.g. to compute the effect of a change of households and/or house characteristics on energy demand.

Step A

The specification of the demand system is based on a-priori economic reasoning, and we do not know if it is appropriate for the data at hand. For instance if

$$H_0 : \alpha_{3i} = 0, \qquad i = e, g$$

then the relative price of energy would not have a direct effect on the two demands. If the `null hypothesis'H_0 were true, any attempt to reduce energy demand (shares) by imposing an energy tax would be vacuous. One hence needs to reject H_0 in order to confidently estimate the parameters α_{3i} that contribute to the so-called *price elasticities* of demand $\partial \ln x_{ih} / \partial \ln p_i$. A similar caveat applies to all coefficients α_{ji} in the demand system.

Testing the `null hypothesis'H_0 within the model can be performed via a t -test. In the paper, the test of hypothesis H_0 for $i = e$ gave a t -test of 8.5, which is significant at the 5% level. The same hypothesis for gas gave a t -test of 4.8, which is also significant at the 5% level. This leads to a rejection of H_0 , and hence supports the existence of a significant relation between $\ln\left(\frac{p^{en}}{p^{ne}}\right)$ and s_{eh} and s_{gh} .

Step B

When step A leads to rejecting the null hypothesis of no influence of some explanatory variable in X on Y , we wish to estimate the α_{ji} coefficients in the demand system. This is accomplished through appropriate estimators. In the case of the paper in question, the estimation was performed with a panel fixed effect model, taking first differences and applying the generalized least squares (GLS) estimator, see e.g. Wooldridge (2011) for details. For cross-section (and panel) data, steps A and B can be performed simultaneously by estimating a model and testing the significance of its

coefficients.

For instance, the α_{3e} coefficient was estimated to be equal to 1.774, with a 95% *confidence interval* equal to (1.367, 2.181), see Table 1 in the paper. This confidence interval provides a measure of the uncertainty associated with the estimation of α_{3e} : the point estimated of 1.77 can be considered as the `most likely' value for α_{3e} on the basis of the sample evidence, and the interval of values (1.367, 2.181) contains α_{3e} with 95% probability.

By comparison, the α_{3g} coefficient was estimated to be equal to 1.318 with a 95% confidence interval equal to (0.785,1.851), see Table 2 in the paper. Hence the point estimate of α_{3g} is smaller that α_{3e} , even though the confidence intervals for the two coefficients overlap: (0.785,1.851) $\cap$ (1.36, 2.18)=(1.36, 1.851) $\neq \emptyset$. This implies that we do not have sufficient statistical evidence to reject the hypothesis that $\alpha_{3g} = \alpha_{3e}$.

These examples illustrate some of the wealth of information that can be produced via appropriate econometric techniques. This information includes test results, point estimates and uncertainty measures associated with the estimation.

Step C

The aim of the paper is to evaluate ex post the effect of a specific energy tax on energy demand in the NL. Here is a quote from the introduction of the paper:

> *In order to reduce energy consumption, some Western governments have introduced energy taxes. In 1996, the Netherlands introduced a Regulating Energy Tax (known as the `REB'). This tax is imposed on final energy consumption of all sectors of the economy. Since the aim is not to raise tax revenues, the revenue of the REB is recycled such that the collective tax burden remains at the pretax level. The general idea behind the regulating energy tax is simple. The additional tax leads to a higher price and as a result, provided there is a downward sloping energy demand curve, to a lesser demand. Since on average households are compensated for their loss of income (by means of the recycling of tax revenues), the tax only changes relative prices. Of course, the success of this policy hinges on the size of the price elasticity, which, as the empirical literature suggests, appears to be rather small.*
> *The objective of this paper is to analyze the success of the above-mentioned policy by looking at the price elasticity of Dutch household energy, in particular of electricity and gas.*

On the basis of the estimates in point B, the authors estimate the elasticity of demand with respect to price, and conclude that in year 1999, for instance, the REB tax induces a 18.9% increase in price of electricity which resulted in a 11% decrease of electricity demand by households. The same calculations performed over the years 1996 to 1999 gave an average increase of electricity price of 13.9% which resulted in an 8% reduction of electricity demand (see their Table 3).

Similar calculations were performed for gas; they gave an average increase of gas price of 16.1% which corresponds to a 4.4% reduction of demand (see their Table 4). They conclude that demand for electricity is much more elastic to price than the one of gas, possibly because gas is the preferred energy for house heating, and which cannot be easily substituted.

Because the demand system contains also information of house and household characteristics, the authors calculated also the effects “relocation” (living in a row house instead of a detached house) “better insulation” (double glass windows), “decrease of family size from 3 to 2” etc. All in all they were able to compare the effects of measures that are potentially viable (like incentivizing the use of double glass) as well as counterfactual measures that are not potentially viable (like house relocation).

Even though the authors do not present this in the paper, they could have presented an uncertainty analysis associated with the effects of these policy changes. This could be presented in the form of confidence intervals for the effects of the energy tax.

This example shows what a demand system is, which data are required for its estimation, the type of inferences that can be made concerning the statistical significance of a relation between X and Y, the type of policy conclusions one can make on the basis of an estimated demand system on cross-section data.

We now turn to time-series data.

3.4 ABC for time-series data – household sector

The same ABC steps can be performed on time-series data. Here we follow the specification in McAvinchey and Yannopoulos (2003) for an energy demand system for heating buildings (with related direct emissions) based on time-series data. This example exemplifies steps A), B) and C) from the Section 3.1. The role played by the type of available data is emphasized.

The purpose of the study was to forecast demand for energy in the UK and Germany (DE), possibly accounting for the interplay between non-stationarity and structural change. The study found that accounting for the interaction of both greatly improved forecasts.

Data

The study was performed on aggregate quarterly time-series for the UK and DE over the period 1978Q1 to 1994Q2. The time dimension is indicated by $t = 1, 2, \ldots$

The available data can be grouped in the following categories:

Energy demands: Let x_{it} indicate the demand for energy type i at time t, where $i = e, o, c, g$ where e stands for electricity, o for oil, c for coal and g for gas. Let E_t stand for total energy demand, and let $s_{it} := x_{it}/E_t$ be the share or energy type i.

Prices: Let P_i indicate price of energy type i

Specification

The specification consists of a level demand system for energy given in eq. (3) of the paper, which is rewritten here as follows:

$$s_{et} = \beta_{e1} + \beta_{e2}\ln\left(\frac{P_{et}}{P_{ct}}\right) + \beta_{e3}\ln\left(\frac{P_{ot}}{P_{ct}}\right) + \beta_{e4}\ln\left(\frac{P_{gt}}{P_{ct}}\right) + \beta_{e5}\ln(E_t) + \beta_{e6}t + z_{et}$$

$$s_{ot} = \beta_{o1} + \beta_{o2}\ln\left(\frac{P_{et}}{P_{ct}}\right) + \beta_{o3}\ln\left(\frac{P_{ot}}{P_{ct}}\right) + \beta_{o4}\ln\left(\frac{P_{gt}}{P_{ct}}\right) + \beta_{o5}\ln(E_t) + \beta_{o6}t + z_{ot}$$

$$s_{gt} = \beta_{g1} + \beta_{g2}\ln\left(\frac{P_{et}}{P_{ct}}\right) + \beta_{g3}\ln\left(\frac{P_{ot}}{P_{ct}}\right) + \beta_{g4}\ln\left(\frac{P_{gt}}{P_{ct}}\right) + \beta_{g5}\ln(E_t) + \beta_{g6}t + z_{gt}$$

The dependent variables of the system are $S_t := (s_{et}, s_{ot}, s_{gt})'$. Note that one of the shares is left out, because of the summing up restriction $\sum_{i\in\{e,o,c,g\}} s_{ih} = 1$. The `explanatory variables' are $X_t := (X_{1t}, X_{2t}, X_{3t}, X_{4t}, X_{5t})' := (\ln(P_{et}/P_{ct}), \ln(P_{ot}/P_{ct}), \ln(P_{gt}/P_{ct}), \ln(E_t), t)'$. Here z_{it} are stationary stochastic errors.

The level demand system is a set of linear combinations of trending (i.e. non-stationary) variables that is stationary; this situation was called *cointegration* by Engle and Granger (1987). The number of (linearly independent) linear combinations in the demand system is called the *cointegration rank*.

The level demand system describes *equilibrium* relations, that are satisfied by the trending variables contained in a system including $(S_t', X_t)'$. Here z_{it} are stationary stochastic errors, unlike Y_t and X_t that are trending. These stationary errors are grouped into the vector $z_t := (z_{et}, z_{ot}, z_{gt})'$, which represents *deviations from the* level demand system. β_{ij} are the coefficients of the level demand system.

This is again a level demand system because:

- it relates energy demand (shares) to prices of the same energy source as well as of alternative energy and of alternative goods; this allows e.g. to compute the effect of a change in relative prices on the different energy demands;
- it relates energy demand (shares) to total energy demand; this allows e.g. to compute the effect of a change in total energy demand.

Note that the system cannot relates energy demand to households and house characteristics, because it uses aggregate data. However, it can be used to forecast energy demand, as it is shown below.

The short run adjustment is given by the *error correction model* in eq. (9) of the paper, which can be written as

$$\Delta s_{et} = c_e + \sum_{i\in\{e,o,g\}}\sum_{\ell=1}^{m}\gamma_{ei,\ell}\Delta s_{i,t-\ell} + \sum_{i=1}^{4}\sum_{\ell=1}^{m}\omega_{ei,\ell}\Delta X_{i,t-\ell} + \sum_{i\in\{e,o,g\}}\alpha_{ei}z_{i,t-1} + \varepsilon_{et}$$

$$\Delta s_{ot} = c_o + \sum_{i\in\{e,o,g\}}\sum_{\ell=1}^{m}\gamma_{oi,\ell}\Delta s_{i,t-\ell} + \sum_{i=1}^{4}\sum_{\ell=1}^{m}\omega_{oi,\ell}\Delta X_{i,t-\ell} + \sum_{i\in\{e,o,g\}}\alpha_{oi}z_{i,t-1} + \varepsilon_{et}$$

$$\Delta s_{gt} = c_g + \sum_{i\in\{e,o,g\}}\sum_{\ell=1}^{m}\gamma_{gi,\ell}\Delta s_{i,t-\ell} + \sum_{i=1}^{4}\sum_{\ell=1}^{m}\omega_{gi,\ell}\Delta X_{i,t-\ell} + \sum_{i\in\{e,o,g\}}\alpha_{gi}z_{i,t-1} + \varepsilon_{et}$$

Here the α_{ji} coefficients describe adjustment to the level shares, as measured by the lagged deviations $z_{i,t-1}$; this mechanism of correction of previous deviations is what was called *error correction* in the late 1970s and early 1980s before cointegration was invented, and it is now called *equilibrium correction* by Hendry.

This system is called the adjustment system, and contains predictive equations. In fact all the explanatory variables that appear on the right-hand-side are lagged, so that they can be used to form prediction for the left-hand-side variables.

Step A

In order to avoid the spurious regression problem, one needs to perform statistical tests in a specific way. In order to see if some β coefficients in the level demand system are different from 0, one needs first to test for the cointegration rank, i.e. the number of level relations. Once the cointegration rank is selected one can test hypothesis like

$$H_0 : \beta_{i3} = 0, \qquad i = e, o, g$$

to see if level demand depends on the price of oil. Several hypotheses of this type can be performed on the β coefficients. Once tests on β are completed, one can test hypothesis on the adjustment coefficients α_{ji} as well as on the γ and ω coefficients. All insignificant coefficients can be constrained to be 0.

Step B

After step A, one can estimate jointly the level-demand system and the adjustment-ECM systems. This is accomplished through appropriate estimators, usually by maximum likelihood. This leads to point-estimates and confidence intervals, similarly to the case of cross-sectional data.

Step C

The aim of the paper was to evaluate the forecast performance of the model. Additionally, one can perform a number of policy simulations, also through the notion of impulse response analysis; see for instance Luetkephol (2005).

This example shows that a demand system can be also estimated on time-series data; for time-series data one has a system in levels and an adjustment system. This gives a forecasting model over time, that can be used to make forecast one or more steps into the future. Note, however, that with time-series data, one cannot measure effects of household- and house- characteristics as in Section 3.2.

Like in the cross sectional case, time-series models can be used to make counterfactual calculations, like calculating the effect of a price increase on energy demand. This calculation can also be made *dynamically*, and this delivers the so-called *impulse responses*.

3.5 ABC for panel data – Transport sector

In this section we present an example of panel-data analysis on the transport sector (TS), which appears of interest for the present project. This is an examples of Panel-data methods, and hence representative of the possible results that can be obtained with these methods. This example also illustrates how the decomposition analysis and the econometric analysis can be fruitfully combined. Here we follow Johansson and Schipper (1997), which is a widely-cited study on fuel demand analysis.

Data

In this study, data regards 12 OECD countries for the years 1973-1992 (namely: USA, UK, Japan, Australia, Germany, France, Italy, the Netherlands, Sweden, Denmark, Norway, Finland). They include car-fuel consumption (differentiated by type: petrol, diesel LPG and CNG), average fuel prices, national income, fuel-intensity standards, taxes on news cars, yearly fees, geographical variables (population density). Many important data-issues are discussed in detail in the study.

Specification

Fuel demand per capita Q is decomposed as the product of vehicle-stock per capita S, fuel intensity I (fuel consumption per km driven) and mean driving distance per car D, where Q=S • I • D. Each of the components S, I, and D can be better modeled separately (and interdependently) than their aggregate. In particular a *recursive* system approach is taken, where D is modelled as a function of S and I (as well as other variables), and I and S are modelled as a function of other variables.

Specifically, let

- S_{it} be car stock in country i and time t,
- P be fuel price,
- Y be national income,
- T be taxation,
- G be population density.

The study specifies the country-homogeneous dynamic relations

$$\begin{aligned}
\ln S_{it} &= \alpha_0 + \alpha_1 \ln S_{i,t-1} + \alpha_2 \ln P_{it} + \alpha_3 \ln Y_{it} + \alpha_4 T_i + \alpha_5 G_i + u_{it}^S, \\
\ln I_{it} &= \beta_0 + \beta_1 \ln I_{i,t-1} + \beta_2 \ln P_{it} + \beta_3 \ln Y_{it} + \beta_4 T_i + \beta_5 G_i + u_{it}^I, \\
\ln D_{it} &= \gamma_0 + \gamma_1 \ln D_{i,t-1} + \gamma_2 \ln(P_{it} I_{it}) + \gamma_3 \ln Y_{it} + \gamma_4 T_i + \gamma_5 G_i + \gamma_6 \ln S_{it} + u_{it}^D.
\end{aligned}$$

Here the u_{it} terms denote stochastic error terms. Observe that $P_{it} I_{it}$ measures the mean cost per kilometre. Rewriting

$$\gamma_2 \ln(P_{it} I_{it}) = \gamma_2 \ln P_{it} + \gamma_2 \ln I_{it}$$

one sees that the specification imposes the same coefficient on two terms, i.e. that the price elasticity of car stock is equal to fuel intensity elasticity. On the latter there exists a quite extensive *rebound effect* literature, where increases in fuel efficiency are not totally transferred to a decrease in fuel demand.

The system of 3 equations is estimated, and steps A) and B) are performed; namely A)significance of all coefficients is tested, and B)all significant coefficients are estimated appropriately.

In order to find the effects of P, Y, T and G on Q, the following steps are performed.

1. The long run responses is calculated for each equation. This step can be described as 'dropping the t subscript' and solving for the dependent variable. This gives (dropping also the error term for simplicity)

$$\begin{aligned}
\ln S_i &= \frac{1}{1-\alpha_1}(\alpha_0 + \alpha_2 \ln P_i + \alpha_3 \ln Y_i + \alpha_4 T_i + \alpha_5 G_i), \\
\ln I_i &= \frac{1}{1-\beta_1}(\beta_0 + \beta_2 \ln P_i + \beta_3 \ln Y_i + \beta_4 T_i + \beta_5 G_i), \\
\ln D_i &= \frac{1}{1-\gamma_1}(\gamma_0 + \gamma_2 \ln P_i + \gamma_2 \ln I_i + \gamma_3 \ln Y_i + \gamma_4 T_i + \gamma_5 G_i + \gamma_6 \ln S_i).
\end{aligned}$$

2. Eliminating the simultaneity of the system; this consists of replacing $\ln S_i$ and $\ln I_i$ in the last equation by the r.h.s. of the first two equations. This gives a system where the third equation is replace by

$$
\begin{aligned}
\ln D_i = & \frac{1}{1-\gamma_1}(\gamma_0 + \frac{\gamma_2}{1-\beta_1}\beta_0 + \frac{\gamma_6}{1-\alpha_1}\alpha_0) + \\
& + \frac{1}{1-\gamma_1}\left(\gamma_2 + \frac{\gamma_2}{1-\beta_1}\beta_2 + \frac{\gamma_6}{1-\alpha_1}\alpha_2\right)\ln P_i + \\
& + \frac{1}{1-\gamma_1}\left(\gamma_3 + \frac{\gamma_2}{1-\beta_1}\beta_3 + \frac{\gamma_6}{1-\alpha_1}\alpha_3\right)\ln Y_i + \\
& + \frac{1}{1-\gamma_1}\left(\gamma_4 + \frac{\gamma_2}{1-\beta_1}\beta_4 + \frac{\gamma_6}{1-\alpha_1}\alpha_4\right)T_i + \\
& + \frac{1}{1-\gamma_1}\left(\gamma_5 + \frac{\gamma_2}{1-\beta_1}\beta_5 + \frac{\gamma_6}{1-\alpha_1}\alpha_5\right)G_i
\end{aligned}
$$

Finally Q is found aggregating $S_i I_i D_i$, i.e.

$$
\begin{aligned}
\ln Q_i = & \ln S_i + \ln I_i + \ln D_i \\
= & \frac{1}{1-\gamma_1}(\gamma_0 + \frac{\gamma_2}{1-\beta_1}\beta_0 + \frac{\gamma_6}{1-\alpha_1}\alpha_0) + \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_0}{1-\beta_1} \\
& + \left(\frac{1}{1-\gamma_1}\left(\gamma_2 + \frac{\gamma_2}{1-\beta_1}\beta_2 + \frac{\gamma_6}{1-\alpha_1}\alpha_2\right) + \frac{\alpha_2}{1-\alpha_1} + \frac{\beta_2}{1-\beta_1}\right)\ln P_i + \\
& + \left(\frac{1}{1-\gamma_1}\left(\gamma_3 + \frac{\gamma_2}{1-\beta_1}\beta_3 + \frac{\gamma_6}{1-\alpha_1}\alpha_3\right) + \frac{\alpha_3}{1-\alpha_1} + \frac{\beta_3}{1-\beta_1}\right)\ln Y_i + \\
& + \left(\frac{1}{1-\gamma_1}\left(\gamma_4 + \frac{\gamma_2}{1-\beta_1}\beta_4 + \frac{\gamma_6}{1-\alpha_1}\alpha_4\right) + \frac{\alpha_4}{1-\alpha_1} + \frac{\beta_4}{1-\beta_1}\right)T_i + \\
& + \left(\frac{1}{1-\gamma_1}\left(\gamma_5 + \frac{\gamma_2}{1-\beta_1}\beta_5 + \frac{\gamma_6}{1-\alpha_1}\alpha_5\right) + \frac{\alpha_5}{1-\alpha_1} + \frac{\beta_5}{1-\beta_1}\right)G_i
\end{aligned}
$$

This formula shows how the different (semi)elasticities in the model are combined to give the effects of prices, income, taxation and population density on fuel demand. We call this the *impact evaluation formula*.[6]

The steps A), B) and C) are qualitatively similar to the ones described in the previous sections. Note that step A) makes sure that each coefficient in the impact evaluation formula is indeed relevant. Remark that the specification adopts a country-homogeneity assumption, which may be a limitation for the validity of the study.

3.6 ABC for time-series data – transport sector

In this section we present one time-series analysis on the TS that appears of interest for the development of the present project. Again, the focus on the A, B and C steps. Here we follow Liddle (2009).

Data

In this study, data refers to the USA for the years 1946-2006. They include: real gross domestic product per capita (GDP), vehicle-miles per capita (VMT), number of registered vehicles per capita (REG), retail gasoline price (PRICE). The data set is of time-series type, as it includes a sufficient number of time periods.

[6] Recall that *elasticities* are percentage change in output (as for Q) for a percentage change in input (as for P or Y)

A *semi-elasticity* (or *semielasticity*) gives instead the percentage change in the left-hand-side Q in terms of a change (not percentage-wise) of the input. This is the case for the coefficients to the variables T and G.

Specification

A cointegration analysis is performed on 4 time-series of the natural logarithm of VMT, PRICE, GDP, REG, here labeled as $x_t := (x_{1t}, x_{2t}, x_{3t}, x_{4t})' = (\log \text{VMT}, \log \text{PRICE}, \log \text{GDP}, \log \text{REG})'$. In the following i indicates one of the variables in x_t; for instance x_{2t} is the second variable in x_t.

The authors argue in favor of a single cointegrating relationship, and they estimate an error correction model of the type

$$\Delta x_{it} = \gamma_{0i} + \sum_{i=1}^{4} \sum_{j=1}^{\ell} \gamma_{ij} \Delta x_{it-j} + \alpha_i ECT_{t-1} + u_{it}$$

for *i*=1,2,3,4 and where ECT denotes the error correction term in the level relation, which is defined in the following level (cointegrating) relation

$$x_{1t} = -0.18x_{2t} + 0.46x_{3t} + 0.62x_{4t} + 4.56 - ECT_t$$

This level relation is interpreted to give level elasticities of fuel demand with respect to prices, income and number of registered vehicles.

The steps A), B) and C) are qualitatively similar to the ones described in the previous sections on cointegration in demand systems.

Note that the specification is country-specific, and that, unlike the paper by Johansson and Schipper (1997) it does not need to assume country homogeneity. Remark that this made possible by the availability of 60 years of data (time-series dataset).

The criticism that the USA in 1946 may be very different from the one in 2006 can be mitigated by the observation that, while the whole economy may have changed dramatically over this period, the assumption of time-homogeneity is only made for economic decisions on fuel demand, for which it could be a reasonable assumption.[7]

3.7 What data can tell

Long time-series data on all EU countries would allow a full time-series analysis for each country. This would permit to distinguish behavior at country level, which appears recommendable both from a modelling perspective (different countries may have different tastes, technologies etc.) and from a policy perspective.

However, one may face lack of appropriately long time-series. In this case one can resort to panel-data methods, if one is willing to make a country-homogeneity assumption. Panel data sets may also include variables that are not measurable or that may be difficult to measure at the aggregated level. It may be possible, for instance, to estimate the effects of changing the insulation of houses on demand for gas, as in Section 3.2. This effect may be hard to estimate on more aggregate data.

The time-homogeneity assumption in time-series models is homologous to the cross-sectional homogeneity assumption, which is used in many cross-sectional or panel data models. One may say

[7] For more on the homogeneity assumption, see the general considerations formulated in the following section.

that all econometric models need to assume some invariance in order to postulate a single model on data collected on different time periods or on different units. This is the cost paid for the construction of an economic model, which is by definition a "simplification of reality", i.e. a simplified framework designed to illustrate complex processes. Quite obviously, the degree of acceptability of (both types of) homogeneity assumptions depends on the data and the model under consideration.

These homogeneity assumptions should not be taken for granted, but rather they should be tested on the data. In the time-series literature there is a well-established tradition (dating back at least to the work of Box and Jenkins, 1970) that prescribes to check the adequacy of the assumptions of the model[8]. This tradition is also present in current time-series econometrics. For instance, it is at the basis of David Hendry's econometric methodology, see Spanos (1989) and Hendry and Nielsen (2007).

This methodology prescribes to fit a general statistical model to the data, such as a Vector Autoregressive (VAR) model. This estimation is followed by a check that the assumptions of the model are consistent with the data; this analysis usually consists of tests on the absence of autocorrelation and on the normality of residuals. This modeling phase is called mis-specification analysis, because it controls if the model is mis-specified. Only when the mis-specification analysis signals that the modeling assumptions are consistent with the data, the model is used to test and estimate economic specifications.

In the present project, we will use state-of-the-art mis-specification tests in line with David Hendry's econometric methodology, in order to check the assumptions of the model, which include the assumption of time-invariance. A exposition of the mis-specification tests that will be employed in this project can be found e.g. in Luetkepohl and Kratzig (2004).

In summary, different types of data can provide answers to different questions. Time-series data are preferable for near term predictions with drivers that are available at an aggregate level, like fuel prices and income. Panel data can help to estimate the effect of other local policies, counting on micro-economic effects for which no aggregate data exist.

[8] Box and Jenkins (1970) suggest to perform a "modeling cycle", that consists of the following steps. A) Determine which model to fit to the data (for the class of stationary ARMA model, this involves the analysis of the data autocorrelations functions). 2) Estimate the chosen model. 3) Check the model fit to find if the modeling assumptions are inconsistent with the data. This last phase consists of the analysis of the autocorrelation function of residuals. If this signals that the assumptions of the model are inconsistent with the data, the cycle is repeated, and a model is used only when phase 3) is verified.

Chapter 4
Expectations and cautions

4.1 Summary

The project aims at understanding the link between changes in drivers and emissions and the ultimate goal is to draw policy-relevant conclusions from this analysis. The literature review on emission trends and drivers provided in this document reveals two common methods of trend analysis:

1. decomposition methods (based on a commonly-agreed paradigm) for describing sources of variation in emission data (DCE)
2. econometric vector autoregressions (VAR) for finding the dynamic link (cause-and-effects) between emissions and their economic, technological, and behavioral drivers.

DCE methods are useful for finding an appropriate disaggregation level, where each decomposed indicator can be dynamically modeled (via econometric methods - VAR) as a function of drivers. The VAR approach is based on cause-effect econometric models that can ascertain the causality of a driver in past trends, and that can measure quantitatively the impact of the driver on emissions within an uncertainty margin. Focusing on the dynamics aspects and looking for cointegration between emission-related and income-related time-series, the second method complements the first.

Given the dynamic importance when assessing the impact of drivers in near-term scenarios, the proposed methodology for this study is hence to use econometric methods as in 2., possibly in combination with decomposition methods in 1. above, using time-series data. The list of decomposed indicators will be complemented with other potential drivers from the econometric literature.

An appropriate level of depth within sectors and even sub-sectors is required for the analysis. The non-ETS sectors, such as road transport (TS) and households (HS), are suitable to illustrate the feasibility of this combined two-step methodology in Member States of the EU-27. Its implementation demands that appropriate time-series data is available.

Cross sectional and panel data analyses are also of potential interest, in order to evaluate alternative drivers that are not available with time-series data, as illustrated by the HS demand system above. This is however beyond the scope of the present project.

4.2 Outlook

Given the limited scope and budget of the project, the need for a country-specific approach and for a subdivision into sectors (or even subsectors), a selection of a pilot case study appears mandatory. The pilot case study will be used to illustrate the potential benefits of the proposed methodology, which will consist of both: DCE methods (as in 1.) and econometric, time-series VAR techniques (as in 2.).

The potential for further policy making is highest in the non-ETS sectors, such as TS and HS. This suggests to consider these sectors (or some of their subsectors) for the pilot study on EU27.

The energy needs in the HS sector are diverse over Europe; they vary widely possibly due to tradition, climate, available resources and infrastructure. This suggests that a single model is probably unfit to cover all EU27 countries.

The passenger transport within the TS in EU27 is relatively homogeneous. This would allow to have the same model applied to all countries. However, data availability is sparse, and it is required to assemble different data sources in order to obtain appropriate time-series. This subsector would yield an interesting application of the cointegration VAR analysis.

The lessons learned from the pilot case study can subsequently be transferred, to the extent possible, to other (sub)sectors. The project aims at consistently using the same methodology for all sectors, but the potential drivers and relationships are expected to be sector-specific and even country-specific in some cases.

References

Albrecht, J., Francois, D., Schoors, K., (2002) A Shapley decomposition of carbon emissions without residuals, Energy Policy 30, pp. 727-736.

Alcantara, V., Duarte, R. (2004) Comparison of energy intensities in European Union countries. Results of a structural decomposition analysis, Energy Policy, 32, pp. 177-189
Ang, B.W. (1999), Is the energy intensity a less useful indicator than the carbon factor in the study of climate change? Energy Policy 27, pp.943–946.

Ang, B.W., Choi, K.-H. (2002) Boundary problem in carbon emission decomposition, Energy Policy, 30, pp. 1201-1205.

Ang, B.W., Liu, F.L., Chew, E.P., (2003) Perfect decomposition techniques in energy and environmental analysis. Energy Policy, 31 (2003) 1561–1566.

Ang, B.W., Liu, N. (2007) Viewpoint: Energy decomposition analysis: IEA model versus other methods, Energy policy 35, pp. 1426-1432.

Ang, B.W., Huang, H.C., Mua, M.R.(2009) Properties and linkages of some index decomposition analysis methods, Energy Policy 37, pp. 4624-4632.

Baek, J., Y. Cho, and W. W. Koo (2009). The environmental consequences of globalization: A country-specific time-series analysis. *Ecological Economics 68* (8-9), 2255–2264.

Berkhout P.H.,A. Ferrer-i-Carbonell, J. C. Muskens (2004) The ex post impact of an energy tax on household energy demand, Energy Economics 26, 297-317.

Box, G. and Jenkins, G. (1970) “Time series analysis: Forecasting and control”, San Francisco: Holden-Day.

Davis L. W. and L. Kilian (2011) Estimating the effect of a gasoline tax on carbon emissions, Journal of Applied Econometrics 26, 1187-1214.

EC (2010) Report from the Commission to the European Parliament and the Council: Progress towards achieving the Kyoto objectives" – COM (2010) 569 final.

EEA (2009) “Transport at a crossroads. TERM 2008: indicators tracking transport and environment in the European Union” – EEA Report No 3/2009

EEA (2010) "Tracking progress towards Kyoto and 2020 targets in Europe" - EEA Report No 7/2010.

EEA (2011) “Greenhouse gas emissions in Europe: a retrospective trend analysis for the period 1990-2008” – EEA Report No 6/2011, ISSN 1725-9177

Engle, R.F. and Granger, C.W.J.(1987) Co-integration and Error Correction: Representation, Estimation, and Testing, Econometrica 55, 251-276.

Engle R.F. and White H. (1999) Cointegration, Causality, and Forecasting, Festschrift in Honour of Clive W. J. Granger, Oxford University Press.

ESTAT (2012) EuroSTAT Structural indicators for Environment, http://epp.eurostat.ec.europa.eu/portal/page/portal/structural_indicators/indicators/environment

Geweke J. (1984) Inference and causality in economic time-series Chapter 19 in Handbook of Econometrics, Volume II, (Griliches and Intriligatoreds), Elsevier Science Puhlishers B V, I984.

Gingrich, W., Kushova, P., Steinberger, J., (2011) Long-term changes in CO2 emissions in Austria and Czechoslovakia—Identifying the drivers of environmental pressures Energy Policy 39(2011), pp. 535-545

Gross, C., (2012), Explaining the (non-) causality between energy and economic growth in the U.S.—A multivariate sectoral analysis, Energy Economics 34 (2012), pp. 489-499

Haavelmo T. (1944) The Probability Approach in Econometrics, Econometrica12, iii-vi+1-115.

Hendry D. F. (1995) Dynamic Econometrics, Oxford University Press.

Hendry D. F. & B. Nielsen (2007) "Econometric Modeling: A Likelihood Approach", Princeton University Press.

IEA/OECD (2007) *Energy Use in the New Millennium: Trends in IEA Countries,* IEA/OECD Paris, http://www.iea.org/w/bookshop/add.aspx?id=312.

IEA/OECD (2007) Tracking Industrial Energy Efficiency and CO2 Emissions, IEA/OECD Paris, http://www.iea.org/textbase/nppdf/free/2007/tracking_emissions.pdf.

IEA/OECD (2004) Oil Crises and Climate Challenges: 30 Years of Energy Use in IEA Countries, IEA/OECD, Paris, www.iea.org/Textbase/publications/free_new_Desc.asp?PUBS_ID=1260.

IMF (2011), World Economic Outlook Database, International Monetary Fund, April 2011.

Jalil, A., Feridun, M. (2011) The impact of growth, energy and financial development on the environment in china: A cointegration analysis, Energy Economics, 33 (2011), pp. 283-291

Jinchao, Z., Kotani, K., (2012), The determinants of household energy demand in rural Beijing: Can environmentally friendly technologies be effective?, Energy Economics 34 (2012), pp. 381-388

Johansson, O. and Schipper, L.(1997) Measuring the long-run fuel demand for cars: separate estimations of vehicle stock, mean fuel consumption per km and mean annual driving distance. Journal of Transport Economics and Policy 31, 277-292.

Johansen, S. (2005) The interpretation of cointegrating coefficients in the cointegrated vector autoregressive model. *Oxford Bulletin of Economics and Statistics 67, 93-104.*

Kaivo-oja, J., Luukkanen, J., (2004), The European Union balancing between CO2 reduction commitments and growth policies: decomposition analyses, Energy Policy 32 (2004), pp. 1511-1530

Karathodorou, N., Graham, D.J., Noland, R.B. (2010), Estimating the effect of urban density on fuel demand, Energy Economics 32 (2010), pp. 86-92

Kaya Y. and Yokobri, K., (1993) Environment, Energy, and Economy: strategies for sustainability. (the output of the Conference on Global Environment, Energy, and Economic Development), (1993) Tokyo, Japan

Lee, C.-C., Chiuen, M.-S. (2010) Dynamic modeling of energy consumption, capital stock, and real income in G-7 countries., Energy Economics, 32 (2010), pp. 564-581

Liddle B. (2009) Long-run relationship among transport demand, income, and gasoline price for the US, Transportation Research Part D 14, 73-82.

Luetkephol H. (2005) A new introduction to multiple time-series analysis, Springer.

Luetkepohl, H. and M. Kratzig (2004) "Applied time series econometrics", Cambridge, Cambridge University Press.

Luukkanen, J. Kaivo-oja, J., (2002), ASEAN tigers and sustainability of energy use: decomposition analysis of energy and CO2 efficiency dynamics, Energy Policy 30 (2002), pp.281-292

McAvinchey, I. D. and A. Yannopoulos (2003) Stationarity, structural change and specification in a demand system - the case of energy, Energy Economics 25, 65-92.

Meehl, G.A., T.F. Stocker, W.D. Collins, P. Friedlingstein, A.T. Gaye, J.M. Gregory, A. Kitoh, R. Knutti, J.M. Murphy, A. Noda, S.C.B. Raper, I.G. Watterson, A.J. Weaver and Z.-C. Zhao, 2007: Global Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, Chapter 10.1

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S.C.B., Frieler, K., Knutti, R., Frame, D.J and M.R. Allen (2009) Greenhouse-gas emission targets for limiting global warming to 2°C, Nature, 458, 1158-1162, doi:10.1038/nature08017

Nakićenović, N., and R. Swart (eds.), 2000: Special Report on Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and NY, 599pp., Chapter 3.1

ODYSSEE-MURE (2010) ODYSSEE database for Energy Efficiency Indicators in Europe, http://www.odyssee-indicators.org/database/database_content.php

Olivier, J.G.J., G. Janssens-Maenhout, J. A.H.W. Peters, J. Wilson (2011) Long-term trend in global CO2 emissions 2011 report,ISBN 978-90-78645-68-9, http://edgar.jrc.ec.europa.eu/news_docs/C02%20Mondiaal_%20webdef_19sept.pdf

Olivier, J.G.J., Janssens-Maenhout, G., Peters, J.A.H.W. (2012), Trends in Global CO2 Emissions: 2012 report EUR 25388, ISBN: 978-92-79-25381-2, http://edgar.jrc.ec.europa.eu/CO2REPORT2012.pdf

Paruolo, P., Murphy, B., Janssens-Maenhout, G. (2011) "Do emissions and income have a common trend? A global time-seriesanalysis 1970-2008." WP Facoltà di Economia 2011/12, Università dell'Insubria, Varese, Italy, availableathttp://eco.uninsubria.it/dipeco/quaderni/files/QF2011_13.pdf

Paul, S., Bhattachary, R.N., (2004) CO2 emission from energy use in India: a decomposition analysis, Energy Policy 32 (2004), pp. 585-593

Pearl, J. (2009) Causality, 2nd Edition, Cambridge University Press.

Scholl, L., Schipper, L., Kiang, N., (1996), CO2 emissions from passenger transport: A comparison of international trends from 1973 to 1992, Energy Policy 24(1) – (1996), pp. 17-30

Song, N. Aguilar, F.X., Shifley, S.R., Goerndt, M.E. (2012), factors affecting wood energy consumption by U.S. households, Energy Economics 34 (2012), pp. 389-397

Soytas, U. and Sari, R., (2003), Energy consumption and GDP: causality relationship in G-7 countries and emerging markets, Energy Economics 25 (2003), pp. 33-37

Spanos, A. (1989) "Statistical Foundations of Econometric Modelling", Cambridge University Press

Stern, D. I. (2004). The rise and fall of the Environmental Kuznets Curve. *World Development 32* (8), 1419–1439.

Storchmann, K., (2005), Long-Run Gasoline demand for passenger cars: the role of income distribution, Energy Economics 27 (2005), pp. 25-58

Sun, J.W. (1998) Changes in energy consumption and energy intensity: a complete decomposition model. Energy Economics 20, 85-100.

UNDP (2010) World Population Prospects (WPP), Revision 2010, Paris.

Yule, G.D. (1926) Why Do We Sometimes Get Nonsense - Correlations between Time-Series? - A Study in Sampling and the Nature of Time-series, Journal of the Royal Statistical Society 89.

Wagner, M. (2008). The carbon Kuznets curve: A cloudy picture emitted by bad econometrics? *Resource and Energy Economics 30* (3), 388–408.

Wooldridge J.(2011) Econometric Analysis of Cross Section and Panel Data, second edition, MIT Press.

World Bank, European Commission, International Monetary Fund, OECD, UN (2009), System of National Accounts 2008, ISBN 978-92-1-161522-7,
http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Annex 1: Metadata of datasources

1. Emissions data

a. UNFCCC/EEA/EuroSTAT collected nationally reported emission data

b. EDGARv4.2 bottom up calculated emission data

The Emissions Database for Global Atmospheric Research – version 4.2, or abbreviated as **EDGARv4.2** is a bottom-up emissions database based on JRC's evaluation of internationally reported activity data (i.e. fuel use, land-use, quantity of industrial products, number of animals), and worldwide consistent assumptions on emission factors associated with these activities for each technology and corrected for end-of-pipe abatement measures. The resulting sector-specific emission trends are publicly available as country totals or on a 0.1°x0.1° grid. Since December 2011, the EDGARv4.2 version of EC-JRC/PBL (2011) can be accessed on http://edgar.jrc.ec.europa.eu/overview.php?v=42 and provides historical emission inventory data from 1970-2008 for all world countries for the greenhouse gases CO2, CH4, N2O, HFCs, PFCs and SF6, the air pollutants CO, NMVOC, NOx, SO2, NH3 and PM10 as particulate matter.

Both datasets are distributed with the proxy data, which are available at the EOLO platform. All the emissions are gridded allocating the proxy datasets to each sector or subsector so that no information is lost, even at highest level of detail. Applying consistently the proxy data allows (1) a geocoverage of all world-countries and (2) representation of single multi-pollutant sources. The temporal distribution uses the proxy data of Table 1, the geospatial distribution those of Table 2. The global gridmaps show all a fine resolution with 0.1ºx0.1º grid cells that are left bottom corner centered and that are used as quality check before release of the dataset.

The EDGARv4.2 is an independent scientific estimate of emissions using consistently for all countries a technology based methodology for the bottom-up calculation. The consistency of the EDGARv4.2 data set is three-fold and concerns the following aspects:
(1) geographical coverage: all world countries are taken into account with the same methodology using standard IPCC (2006) emission factors to the extent possible;
(2) sector-specific activities: energy-related as well as agriculture-related activities, covering the range of IPCC categories are for all countries taken into account with the same sector-specific definitions. Agriculture related activities become important when considering all GHGs with the GWP100 metric, as they contributed 12% of the total in 2008. (In the case of CO2, agricultural activities contributed only 0.4% to net total global emissions, as short cycle carbon emissions from biomass and agricultural waste burning were not included, in accordance with UNFCCC emissions inventorying procedures);
(3) chemical substances: the emissions of multiple GHG and air pollutants by a single human activity are modelled with multipollutant single sources in EDGARv4.2 in space and time, which is in particular critical for point sources.
An overview of the main data sources used for the activity and emission factor data of EDGARv4.2 is given in Table 3.

Table 3: Main overview of major data sources used for the activity data, technology mix, emission factor and end-of-pipe abatement in EDGARv4.2 (and EDGARv4.1)

	Energy/transport	Process industry/waste	agriculture
Activity data	IEA statistics for fuel production & combustion (64x94), CDIAC (fossil fuel fires) & World Bank satellite data (venting/flaring) scientific literature	USGS statistics, UN commodity statistics	FAO data (livestock, rice, crop production), IFA statistics on N-fertilisers, IPCC factors, soil type maps, GFED (savannah burning)
Technologies	UCI Platts Power Plant database, mining, transformation & transmission, national fleet distribution	Technology mix for production techniques, waste water treatment	IRRI cultivation ecosystems, manure management systems
Emission abatement	CH4 recovery (underground mining), EURO classes, US Tier classes for vehicles	CH4 recovery, SF6 recovery, urea	
Emission factors	IPCC Guidelines 2006 & 1996, scientific literature, EMEP/EEA Guidebook 2009	IPCC 2006 guidelines, FOD model for landfills	IPCC 2006 guidelines, scientific literature

A first verification of the EDGARv4.2 emission calculation is done with CO2 and the greenhouse gases in the EU-27, because of the availability of detailed UNFCCC data in CRF, with very satisfying results, that Fig.3 illustrates for CO2 and CO2 equivalent. For all EU-27 member states the difference between UNFCCC/EEA/EuroSTAT data and EDGARv4.2 is <3% for CO2 and is<4% for CO2eq.

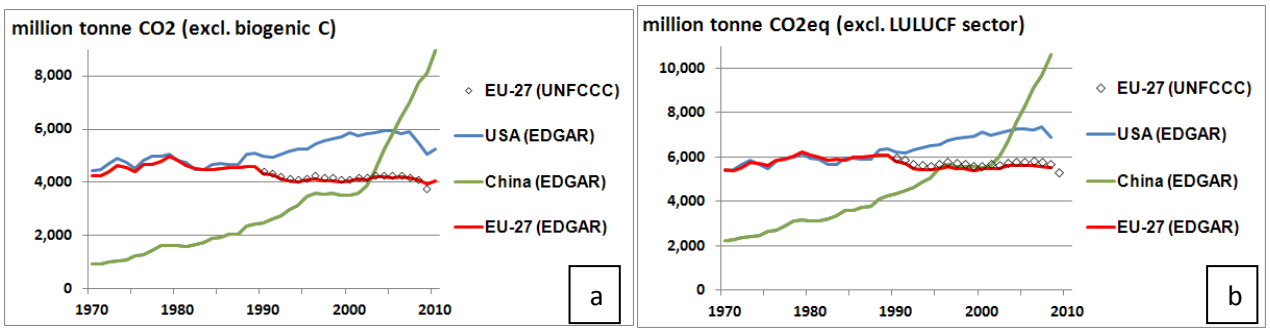


Fig. 3 : CO2 (a) and greenhouse gas (b) emissions time-series 1970-2008 of EDGARv4.2, with fast track extension to 2010 following the same scaling methodology for 2009 and 2010 as performed in the PBL-JRC report of Olivier et al (2011). The CO2 represents all IPCC categories, except the Land-use, Land-use change and forestry sector and includes only long cycle carbon. The short cycle carbon emissions from biomass burning are conform UNFCCC definitions excluded.

2. Economic data

a. Income IMF data

b. PPP data from Eurostat or OECD

c. Income Penn World Table

[The focus of this project is clearly on understanding the drivers for emission trends in EU countries. There is then no specific need to look at the situation in third countries except if this would bring particularly insightful information.]

The IMF's World Economic Outlook database contains selected macroeconomic data series from the statistical appendix of the World Economic Outlook report (September 2011) (http://www.imf.org/external/ns/cs.aspx?id=29), which presents the IMF's analysis and projections of economic developments at the global level, in major countries or country groups. http://www.google.com/publicdata/overview?ds=k3s92bru78li6_Data are available from 1980 to the present, and projections are given for the next two years. Additionally, medium-term projections are available for selected indicators. For some countries, data are incomplete or unavailable for certain years. The data were downloaded from http://www.imf.org/external/pubs/ft/weo/2011/02/weodata/download.aspx on 18th January 2012. The OECD publication "Revenue Statistics 1965-2010: 2011 Edition" presents a unique set of detailed and internationally comparable tax revenue data in a common format for all OECD member countries from 1965 onwards.

Since 1995 Eurostat provides for the European Union Member States the Purchasing power parities (PPPs), as indicators of price level differences across countries and can be used as currency conversion rates to convert expenditures expressed in national currencies into an artificial common currency (the Purchasing Power Standard, PPS), eliminating the effect of price level differences across countries. The main use of PPPs is to convert national accounts aggregates, like the Gross Domestic Product (GDP) of different countries, into comparable volume aggregates. PPPs are also applied in analyses of relative price levels across countries. For this purpose, the PPPs are divided by the current nominal exchange rate to obtain a price level index (PLI) which expresses the price level of a given country relative to another, or relative to a group of countries like the EU27. The production of PPPs is a multilateral exercise involving the National Statistical Institutes of the participating countries, Eurostat and the OECD.

A longer time-series of income for all world countries is given by the Penn World Table estimates, subject of a given uncertainty that is related with the large temporal and geospatial coverage. The Penn World Table, version 7.0 (PWT 7.0), see Heston et al. (2011), using series rgdpl (PPP Converted GDP Per Capita - Laspeyres, derived from growth rates of c, g, i, at 2005 constant prices) for the period 1970-2010. The data were downloaded from http://pwt.econ.upenn.edu/php_site/pwt70/pwt70_form.phpon 12/7/2010.

3. Population data

a. UNDP data or population data from UNSTAT

b. CIESIN data

c. Population data of Eurostat or OECD (and used by IMF)

UNDP reports since 1960 on the demography for all world countries till 2010 in the UNPD (2010) World Population Prospects (WPP), Revision 2010, Paris. The data were downloaded from http://hdrstats.undp.org/en/tables/ on 10/06/2011.

The population from Eurostat are also delivered to UNSTAT. The data were downloaded from http://unstats.un.org/unsd/snaama/selbasicFast.asp on 12/7/2010.

CIESIN providesgridmaps that renders global population data at the scale and extent required to demonstrate the spatial relationship of human populations and the environment across the globe.The purpose of GPW is to provide a spatially disaggregated population layer that is compatible with data sets from social, economic, and Earth science fields. The gridded data set is constructed from national or subnational input units (usually administrative units) of varying resolutions. The native grid cell resolution is 2.5 arc-minutes, or ~5km at the equator, although aggregates at coarser

resolutions are also provided. Separate grids are available for population count and density per grid cell. Population data estimates are provided for 1990, 1995, and 2000, and projected to 2005, 2010, and 2015. The projected grids were produced in collaboration with the United Nations Food and Agriculture Programme (FAO) as Population Count and Density Grid Future Estimates. There is also an extensive map collection that includes population density and sub-national administrative boundary maps (depicting the input units) at country, continental, and global levels.

The demography data are provided on a 2.5minx2.5 min geospatial resolution and therefore were used for population proxy datasets to EDGARv4. It was needed to regridto 0.1degx0.1deg and to multiply with the percentage of cells belonging to a given country. Moreover a refined splitup between urban and rural population was also elaborated. The latest download from http://sedac.ciesin.columbia.edu/gpw/global.jsp dates from 21 September 2011.

4. Road transport data

The country-specific fleet distribution (in percentage)and their respective technologies are given by the international statistics from the International Road Federation (IRF,1990, 2005, 2007) from 1990 onwards, and were analysed in the framework of the EU funded project 'Quantify' (e.g. number of driven vehicle kilometers) and coupled with the respective fuel type. In the project 'Quantify' they took stock figures as a starting point, assumed reasonable annual vehicle mileage, multiplied with average fuel consumption and triangulated everything with the national total fuel consumption until it became reasonable. As a consequence the mileage distribution between different vehicle categories using the same fuel is constrained to about 10-15%.

Additional national distributions were considered for:

- the United States of America and Canada (National Inventory Report, 2008),
- Germany (Statistische Mitteilungen des Kraftfahrt-Bundesamtes, 2007, Bestand von 1950 bis 2008 nach Fahrzeugklassen, http://www.kba.de),
- Japan (Japan Statistical Yearbook,1980 – 2006, Statistic Bureau, Director-General for Policy Planning (Statistical Standards)) & Statistical Research Training Institute, http://stat.gov.jp/english/data),
- Thailand (The Department of Highways, Ministry of Transport and Communications, TABLE 5.8 NUMBER OF VEHICLES REGISTERED BY TYPE: 1997 - 1999)
- Taiwan (Statistical Yearbook of the Republic of China (Taiwan) 2006, edited 2007, Table 128),
- India (Automobile India.com, Two-wheeler statistics, motorcycle statistic 2007 – 2008, http://www. Automobileindia.com) and
- Singapore (Yearbook of Statistic Singapore Transport and Communications, 1996 - 2006).
- For Latin American countries the technology percentage differs due to their different travel demands, with travel distance per year and consumption per travelled kilometer. The technological end-of-pipe progress is reported a World Bank study (World Bank, 1996). The share of non-controlled (NOC) vehicles is published by the Asociacion de fabricantes de automobiles de la Rep. Argentina (ADEFA,2006) regarding vehicles age in Latin American countries. The VKT (Vehicles kilometres travelled) amount depends not only on the category but also on travel demand and population habits, therefore modifications for Argentina, Brazil and Chile need to be taken into account, as indicated by D'Angiola et al (2010) for the MABA .Mexican VKT are reported by Schifter et al, 2004 for Mexico DF. The averaged fuel consumption per kilometer can be taken from the road transport model COPERT IV.

Annex 2: List of potential drivers per sector

Table 4: List of potential drivers from decomposition analyses

SNAP code	IPCC code NFR08	Sector	Decomposed indicator EU-27 (cfr. EEA, 2010, 2011)	Decomposed indicator OECD (cfr. IEA, 2004, 2007) (Ang et al., 1999, 2002, 2003, 2007, 2009)	Decomposed indicator country-specific (cfr. ODYSSEE-MURE, 2010; Gingrich et al, 2011)	Decomposed indicator (cfr. ESTAT, 2010)	(Sun, 1998; Kaivo-oja et al, 2004; Alcantara, 2004; Luukkanen et al, 2002; Paul et al, 2004)
010000	1A1a-1A1b-1A1c	Power industry & fossil fuel manufacturing	Electricity demand (GWh), Share of electricy generation by autoproducers (GWh/GWh), Share of nuclear energy (GWh/GWh), share of renewable energy sources for electricity generation (GWh/GWh), share of heat producton in thermal electricity (GWh/GWh), fuel efficiency of power plants (TJ/GWh), share of biomass for public electricity (TJ/TJ), carbon intensity of thermal public electricity (Mt CO2/TJ)	GDP, Energy services, actual energy use, activity, structure, fuel mix, carbon intensity, energy savings	Primary energy intensity, final energy intensity, energy efficiency index, CO2 intensity, population effect, income effect, emission-coefficient effect, fuel-share effect.	combined heat and power generation, implicit tax rate on energy, electricity generated from renewable sources	Total Primary Energy Supply (TPES), Energy intensity as TPES/ GDP, activity describing the effect of economic growth on sectoral energy use, intensity effect describing the impact of technology and production changes on energy consumption, structural effect describing impact of sectoral share change on energy consumption
020100-020300	1A4	Heating of buildings in commercial & institutional sector		energy per floor area, floor area per employee, employment per value-added	Energy intensity, electric intensity, specific consumption per employee and floor area, CO2 intensity		Total Primary Energy Supply (TPES), Energy intensity as TPES/ GDP, activity describing the effect of economic growth on sectoral energy use, intensity effect describing the impact of technology and production changes on energy consumption, structural effect describing impact of sectoral share change on energy consumption
020200	1A4	Heating of buildings in residential sector	population (million), number of persons per household (million/million), energy use per household (PJ/million), share of electricity and district heat in households' energy use (PJ/PJ), share of biomass in direct fule combustion by households (PJ/PJ), carbon intensity of direct fuel combustion by households (Mt CO2/PJ), temperature (deg. C)	household area per capita, energy use per capita, population, structural composition of energy use (space heating, appliances, water heating, lighting, cooking), energy efficiency, household energy expenditures in total consumption expenditure, energy intensity, energy savings	Energy efficiency index, specific consumption by dwelling and by end uses and by equipment, CO2 intensity, population effect, income effect, emission-coefficient effect, fuel-share effect.		

SNAP code	IPCC code NFR08	Sector	Decomposed indicator EU-27 (cfr. EEA, 2010, 2011)	Decomposed indicator OECD (cfr. IEA, 2004, 2007) (Ang et al., 1999, 2002, 2003, 2007, 2009)	Decomposed indicator country-specific (cfr. ODYSSEE-MURE, 2010; Gingrich et al, 2011)	Decomposed indicator (cfr. ESTAT, 2010)	Decomposed indicator globally (Sun, 1998; Kaivo-oja et al, 2004; Alcantara, 2004; Luukkanen et al, 2002; Paul et al, 2004)
030000-040000	1A2-2	Industrial combustion and processes (heavy manufacturing, metal, mineral, chemical industry - non power)	gross value added (EUR million) for the different industries, energy intensity of manufacturing (TJ/EUR million), share of public electricity in industry's energy use (TJ/TJ), share of fossil fuels combusted by the industries (TJ/TJ), carbon intensity of fuel combused by the industries (Mt CO2/TJ)	manufacturing value-added per capita, raw materials share of manufacturing value-added, energy intensity, structural composition of manufacturing industry	Energy efficiency index, energy intensity by branch, energy intensity at adjusted structure, specific consumption by intensive products (toe/ton), CO2 intensity by sector	energy intensity of economy, resource productivity	Total Primary Energy Supply (TPES), Energy intensity as TPES/GDP, activity describing the effect of economic growth on sectoral energy use, intensity effect describing the impact of technology and production changes on energy consumption, structural effect describing impact of sectoral share change on energy consumption
050000	1B	Fugitive emissions from fossil fuel extraction					
060000	3	Solvent industry and paint application		services value-added per capita			Total Primary Energy Supply (TPES), Energy intensity as TPES/GDP, activity describing the effect of economic growth on sectoral energy use, intensity effect describing the impact of technology and production changes on energy consumption, structural effect describing impact of sectoral share change on energy consumption
070100	1A3bi	road transportation by car	passenger transport demand (Gpkm), share of road transport in total passenger transport (Gpkm/Gpkm), share of private cars in total road passenger transport (Gpkm/Gpkm), fuel intensity of road passenger transport (ktoe/Gpkm), share of biofuels in private cars (ktoe/ktoe), carbon intensity of fossil fuesl use by private cars (Mt CO2/ktoe)	passenger-km per capita, share of passenger travel by mode, car ownership per capita, fuel costs per vehicle-km for cars, fuel use per capita, fuel price, energy per passenger-km by mode, fuel intensity of the car stock, average weight of passenger car, energy savings	energy efficiency index, specific consumption by vehicle in liters/100km, share of CO2 emissions by mode and vehicle	volume of passenger transport relative to GDP, car share of inland passenger transport	
070200-070300-070400-070500	1A3bii-1A3biii-1A3biv	road transportation by vehicles other than cars	freight transport demand (Gtkm), share of road transport in total freight transport (Gtkm/Gtkm), fuel intensity of road freight transport (ktoe/Gtkm), share of biofuels in road freight transport (ktoe/ktoe), carbon intensity of fossil fuels use by road freight transport (Mt CO2/ktoe)	total freight tonne-km per capita, freight transport energy use by mode, energy intensity for trucks, truck average load per vehicle, fuel mix, carbon intensity, energy savings	energy efficiency index, specific consumption by vehicle in liters/100km, share of CO2 emissions by mode and vehicle	volume of freight transport relative to GDP, road share of inland freight transport,	

SNAP code	IPCC code NFR08	Sector	Decomposed indicator EU-27 (cfr. EEA, 2010, 2011)	Decomposed indicator OECD (cfr. IEA, 2004, 2007) (Ang et al., 1999, 2002, 2003, 2007, 2009)	Decomposed indicator country-specific (cfr. ODYSSEE-MURE, 2010; Gingrich et al, 2011)	Decomposed indicator (cfr. ESTAT, 2010)	Decomposed indicator globally (Sun, 1998; Kaivo-oja et al, 2004; Alcantara, 2004; Luukkanen et al, 2002; Paul et al, 2004)
080000	1A3a-1A3c-1A3d-1A3e	non-road mobile sources					
090000	6	waste incineration/disposal & waste water treatment (*)	population (heads), municipal solid waste (MSW) per capita (kt /head), share of MSW incinerated (kt/kt), share of MSW landfilled (kt/kt), CH4 intensity of landfills (Mt CH4/kt), CH4 recovery in landfills (Mt/Mt)			municipal waste generation and treatment, by type of treatment method	
100400	4A	agriculture: enteric fermentation	milk production (T), milk yield (head/T), emission intensity of dairy cattle (Mt CH4/head), number of non-dairy cattle (head), emission intensity of non-dairy cattle (Mt CH4/head)				
100100-100200-100500-100600-100900	4B-4C-4D	agriculture: agricultural soils and manure management	cropland area (kha), animal manure per cropland area (kt N/kha), synthetic fertiliser per cropland area(kt N/kha), other fertiliser per cropland area (kt N/kha), N2O emissions per fertiliser applied (kt N2O/kha)			farmland bird index	
110000+A7	5	land use, land-use change & forestry					
Notes	(*)	The waste sector is defined wih only waste that is not incinerated for energy recovery. The latter is taken up in the power industry (1)					

Table 5: List of potential drivers from econometric analyses

SNAP code	IPCC code NFR08	Sector	Econometrics variables (energy-consumption focused) (Lee et al, 2010; Jalil et al, 2011, Gross, 2012, Soytas, 2003)	Econometrics variables (transport sector) (Johansson & Schipper, 1997; Karathodorou et al, 2010; Storchmann, 2005, Liddle, 2009, Scholl et al, 1996, Davis & Kilian 2011)	Econometrics variables (household sector) (Berkhout et al. 2004; Song et al. 2012, Jingchao et al. 2012, McAvinchey & Yannopoulos 2003)
010000	1A1a-1A1b-1A1c	Power industry & fossil fuel manufacturing	commercial energy use per capita, per capita real GDP, financial development with liquid liabilities to GDP ratio, foreign trade openness ratio, gross value added for power sector, final energy consumption, energy price, capital to energy ratio		
020100-020300	1A4	Heating of buildings in commercial & institutional sector	commercial energy use per capita,financial development with liquid liabilities to GDP ratio, foreign trade openness ratio, net real capital stock, employment as labor variable, and total energy consumption as energy variable, energy price		
020200	1A4	Heating of buildings in residential sector	commercial energy use per capita,financial development with liquid liabilities to GDP ratio, foreign trade openness ratio, net real capital stock, employment as labor variable, and total energy consumption as energy variable		household income, expenditure for energy goods, price for non-energy goods, price for energy goods, general price level corrected with consumer price index, price for electricity including the regulating energy tax, price for gas including the regulating energy tax, household wood energy consumption, price of non-wood energy, number of household members, age of head of household, square meter house area, annual sum of heating degree days, year of construction of the house, urban density, share of the different heating technologies used in the households, temperature, ownership of durable goods (dishwasher etc), house type and age, type of insulation, household education, Energy taxes, geographical location (mountain area)
030000-040000	1A2-2	Industrial combustion and processes (heavy manufacturing, metal, mineral, chemical industry - non power)	commercial energy use per capita,financial development with liquid liabilities to GDP ratio, foreign trade openness ratio, net real capital stock, employment as labor variable, and total energy consumption as energy variable, gross value added for specific industry sector, energy price, capital to energy ratio		
050000	1B	Fugitive emissions from fossil fuel extraction			

SNAP code	IPCC code NFR08	Sector	Econometrics variables (energy-consumption focused) (Lee et al, 2010; Jalil et al, 2011, Gross, 2012, Soytas, 2003)	Econometrics variables (transport sector) (Johansson & Schipper, 1997; Karathodorou et al, 2010; Storchmann, 2005, Liddle, 2009, Scholl et al, 1996, Davis & Kilian 2011)	Econometrics variables (household sector) (Berkhout et al. 2004; Song et al. 2012, Jingchao et al. 2012, McAvinchey & Yannopoulos 2003)
060000	3	Solvent industry and paint application			
070100	1A3bi	road transportation by car	net real capital stock, employment as labor variable, and total energy consumption as energy variable, energy price	car stock per capita, fuel consumption per km, fuel demand per capita, fuel efficiency of the car stock, gasoline intensity, gasoline consumption per car annual distance driven per car and per year, urban density, retail gasoline price, GDP per capita, number of registered vehicles per capita, fuel mix, fuel tax	
070200-070300-070400-070500	1A3bii-1A3biii-1A3biv	road transportation by vehicles other than cars	net real capital stock, employment as labor variable, and total energy consumption as energy variable, energy price	retail gasoline price, GDP per capita, motor fuel use per capita, vehicle miles per capita, number of registered vehicles per capita, fuel mix, modal structure, fuel tax	
080000	1A3a-1A3c-1A3d-1A3e	non-road mobile sources		retail gasoline price, GDP per capita, motor fuel use per capita, vehicle miles per capita, number of registered vehicles per capita, fuel mix, fuel tax	
090000	6	waste incineration/disposal & waste water treatment (*)			
100400	4A	agriculture: enteric fermentation			
100100-100200-100500-100600-100900	4B-4C-4D	agriculture: agricultural soils and manure management			
110000+A7	5	land use, land-use change & forestry			
Notes	(*)	The waste sector is defined wih only waste that is not incinerated for energy recovery. The latter is taken up in the power industry (1)			

European Commission
EUR 25814 EN – Joint Research Centre – Institute for Environment and Sustainability

Title: Analysis of Greenhouse Gas Emission Trends and Drivers: A Survey of Techniques for Emission Decomposition and Econometric Trend analysis

Author(s): Greet Janssens-Maenhout, Paolo Paruolo, Simone Martelli

Luxembourg: Publications Office of the European Union

2013 – 45 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-28602-5 (pdf)
ISBN 978-92-79-28601-8 (cd-rom/dvd)

doi: 10.2788/84075 (online)

Abstract
This report was drafted for the Administrative Arrangement of JRC and DG CLIMA on "Analysis of greenhouse gas emission trends and drivers" and contains the literature review on emission trends and drivers, and describes two common methods of trend analysis: (i) decomposition methods for defining decomposed indicators, (ii) econometric techniques for finding cause - and - effects links and quantitative relationships between emissions and drivers.
Given the dynamic importance when assessing the impact of drivers on emissions in near - term scenarios, the methodology proposed in this report is to use econometric techniques as in (ii), possibly in combination with decomposition methods in (i) using time - series data. Concerning data, an appropriate level of disaggregation within sectors and even sub - sectors is required for the analysis. The potential for future policy making is highest in the non - ETS sectors, such as road transport and households sectors.
The application of the methodology demands that appropriate time - series data is available for the EU - 27 Member states. Such data needs to be carefully assembled from different sources. The availability of data for EU countries appears significantly lower than for the United States, where data of good quality is abundant.
Given the scope of the present project, a selection of a pilot case study appears mandatory. The pilot case study will be used to illustrate the potential benefits of the proposed methodology, which will involve econometric methods (as in ii) possibly combined with decomposition methods (as in i).

LB-NA-25814-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-28602-5

9 789279 286025