

**Sustainability and Agri-Environmental Policy  
in the European Union:  
A Meta-Analytic Investigation**

Katrin Oltmer\*\*  
Peter Nijkamp\*  
Raymond Florax\*\*\*  
Floor Brouwer\*\*

\*VU University Amsterdam, Faculty of Economics and Business Administration,  
Department of Spatial Economics, MASTER-POINT, De Boelelaan 1105, 1081 HV  
Amsterdam, The Netherlands.

\*\*Agricultural Economic Research Institute (LEI), P.O. Box 29703, 2502 LS The  
Hague, The Netherlands.

\*\*\* Purdue University, Department of Agricultural Economics, 403 W. State Street,  
West Lafayette, IN 47907-2056, USA

pn118korffb

## 1. Setting the Scene

Land use and land cover have in recent years become major policy and research issues. In particular as a result of deregulation and decentralisation trends, the number of stakeholders involved with land use planning has increased, while also the economic interests in land has risen. Consequently, land plays a critical role not only in urban rehabilitation projects, real estate development and industrial site planning, but also in environmental management and agricultural land use policy, which is the focus of this paper.

In the history of economic thought, varying attention has been given to land as an economic production factor. A dominant role, for instance, was assigned to land as a basic input to the creation of economic welfare in the period of the physiocrats. In the neoclassical world, land assumed mainly a functional economic position, as productivity and welfare differences between regions could be explained *inter alia* by different soil conditions (see also Giaoutzi and Nijkamp 1994). More recently – partly as a result of the emergence of ecological economics – it has been recognised that land has not only a productive, but also a consumptive meaning (e.g., as recreational resource) in a sustainable development perspective (see for a review Van den Bergh 1996). Furthermore, it is increasingly recognised that the condition of the soil has a variety of direct and indirect impacts on the quality and resilience of ecosystems with serious consequences for biodiversity, not only locally but also globally.

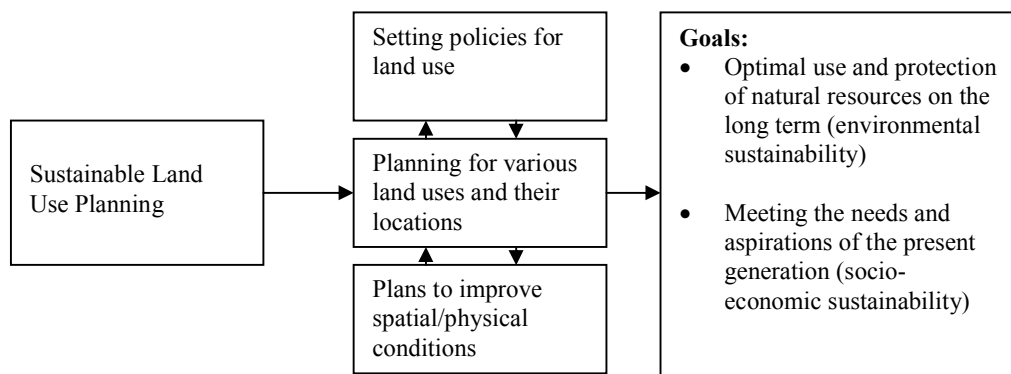
In the spirit of the debate on sustainable development, land use change has recently become a new focal point of interest of both scientists and policy-makers, e.g. in relation to deforestation (Chomitz and Gray 1996), soil rehabilitation (Beinat and Nijkamp 1999, Nijkamp 2000), or urban renewal (Finco and Nijkamp 2000). Clearly, land use is a multi-faceted phenomenon, driven by several economic, demographic, technological and physical factors (such as crop prices, population growth, harvest techniques, climatological factors etc.). In the present paper we will mainly focus on land use in rural and agricultural areas. It is conceivable that a great variety of modelling approaches has been developed to investigate the dynamics in agricultural land use. In a recent study (Groeneveld and Van Ierland 2000), the authors distinguish and review the following types of modelling approaches: analytical models, optimisation models, general equilibrium models, spatial equilibrium models, econometric models, heuristic decision models, empirical models and discrete choice models.

It goes without saying that the sustainability debate has prompted new challenges and research directions in agricultural land use research. For economists, the notion of sustainable development has meant a new major challenge, as they were forced to broaden conventional land use frameworks towards the domain of ecological systems or even international negotiation tables (see for a review again Van den Bergh 1996). Up till now several economic studies are still rather abstract and theoretical in nature (e.g., by seeking for optimal trajectories or game-theoretic equilibria), but an increasing number of studies can be found which offer interesting applied work in the area of agricultural land use (see e.g. Miller and Plantinga 1999 and Parks and Schorr 1997).

The focus on local land use and sustainability conditions has also led to a rising interest in research which moves away from global sustainability analyses towards empirical policy-relevant research at the *regional* land use level (see e.g. Giaoutzi and Nijkamp 1994). This new interest in regional sustainability analysis is caused by several factors: a region is a properly demarcated area with some degree of homogeneity which allows for a more operational empirical investigation; a region is

usually also subject to a properly regulated administrative competence and institutional control, so that there is more scope for a relevant policy analysis of sustainability issues; and finally, the statistical data base at a regional level is often more appropriate for monitoring, analysing and modelling the economy and ecology of an area. In recent international agreements and sustainability studies, the region assumes indeed a prominent position.

In Agenda 21, agreed upon at the Rio Summit, it was stated that land use planning should strive for “*promoting sustainable human settlement development*”. The fulfilment of such a task requires a clear analytical framework. In Figure 1 an illustrative presentation of the scope of sustainable land use planning is given. This figure clarifies that it is no surprise that in recent years land use has increasingly become a battlefield of conflicting interests (see also Frederick and Rosenberg 1994). Over the last centuries, a significant and progressive transformation of natural areas into areas that support agricultural, urban or industrial functions has been observed. Apart from Europe, where both forests and grasslands show a slight expansion, the overall trend is towards a substantial loss of natural land in favour of cropland. The combined pressure of key factors such as population growth, food production, wood production and land tenure arrangements (Pearce 1991) has influenced as much as forty percent of the forests and grasslands of some areas. This trend will continue in the future, as the demand for space and natural resources will probably continue to rise. Irrigated land, cropland, rangeland and pasture will increase in absolute terms, but their availability per capita will also decrease. Without countermeasures, this will necessarily lead to further pressure on land, to an increasing load on environmental quality, and to an impoverishment of natural resource capital. The negative effects of land-use exploitation are manifested in soil erosion, loss of habitats, increased vulnerability of the soil, a decrease in the carrying capacity of land, landscape modification and loss of natural amenities (see Beinat and Nijkamp 1999). Therefore, it is no surprise that in the past years a variety of agricultural land use policies has been developed with the aim to find a balance between economic efficiency and ecological quality.



**Figure 1:** Sustainable planning of land use  
 Source: Van Lier et al. (1994)

Against the above background observations, the present paper has the aim to offer a framework for comparative analysis of agricultural land use practices in various European countries. In this paper, agricultural land use practices are represented by so-called environmental driving force indicators, namely the use of nitrogen fertiliser, livestock density and grassland area. The main emphasis will be on the identification of drivers in agricultural land use practices by means of meta-analytic methods. Some of these drivers may stem from specific policy measures and others from general market or external conditions. Here, drivers stemming from specific policy measures and from the structure of the agricultural sector will be investigated.

This paper considers two major research questions on environmental aspects of agricultural land use. One is concerned with the assessment of environmental effectiveness of agri-environmental policies in the European Union, as a consequence of recent changes in agricultural and environmental policy. From the perspective and the need to draw lessons from comparative case study research in this field, the second research task of this paper deals with a related methodological issues namely whether meta-analysis is a suitable tool for policy assessment of agri-environmental initiatives in the EU.

The various experiences on agricultural policy in various European countries call for a systematic research synthesis and comparison. From a methodological perspective we will deploy here an approach for comparative case study research, named meta-analysis. Meta-analysis has become an established technique in the medical and natural sciences, especially in case of comparative analysis of (semi-) controlled experiments (see e.g. Glass et al. 1984, Hedges and Olkin 1998 and Petitti 1994). Later on it was also extensively used in the social sciences, in particular in experimental psychology, pedagogy, sociology and more recently also in economics (see Matarazzo and Nijkamp 1997; Baaijens and Nijkamp 2000). Meta-analysis aims to synthesise previous research findings or case study results with a view to identifying commonalities that might lend themselves for transferability to other, as yet unexplored cases. The statistics of meta-analysis is in the mean time rather well developed. Especially in case of quantitative case study results a significant progress has been made. In this paper, we will address in particular an ANOVA-type of meta-analysis adapted to effect size estimations to identify common drivers of agricultural dynamics in Europe.

The paper is organised as follows. Section 2 describes agri-environmental policies in the EU, in particular, the structure of the agri-environmental Regulation 2078/92 of the Common Agricultural Policy. Section 3 gives a short introduction into the use of environmental indicators in policy analysis and explains the environmental indicators used in our analysis. Section 4 presents the input data for the analysis that originate from case studies of an EU project. The methodology of meta-analysis and the statistical procedures applied in our analysis are demonstrated in Section 5. Then, Section 6 reports the results, while finally Section 7 gives conclusions and recommendations.

## **2. Agri-Environmental Policies in the European Union**

Along with the MacSharry reform of the European Union's Common Agricultural Policy (CAP) in 1992, three accompanying measures for stimulating the restructuring of the agricultural sector were introduced. One of them is called the agri-environmental measure or, in formal terms, EC-Council Regulation 2078/92. This regulation is concerned with the implementation of special programs that mean to

support and encourage farmers to introduce or continue with agricultural production methods consistent with the requirements of environmental protection and maintenance of characteristic landscapes and the countryside. This implies that it is not only a framework for the stimulation of sustainable agriculture, but also for the multifunctionality and originality of rural space in Europe (Buller 2000). The other two accompanying measures are the early retirement scheme for farmers (Regulation 2079/92) and the afforestation programme of agricultural land (Regulation 2080/92) (see Soufi and Tuddenham 1995; Brouwer and Van Berkum 1996; Buller 2000).

The agri-environmental Regulation 2078/92 is a co-financed instrument. This means that the Member States can apply for co-funding of up to 50% and even up to 75% for Objective 1 regions<sup>1</sup>. The financial source for the accompanying measures is the EAGGF (European Agricultural Guidance and Guarantee Fund) and the amount of money spent on them is rapidly increasing compared to traditional EAGGF expenditures such as the crop or the animal sector. However, the amount of money the EU spends on agri-environmental measures is still not higher than 4% of total CAP expenditures (Buller 2000). The total amount of money spent on agri-environmental measures is obviously higher because, as mentioned above, the national governments also contribute to the financing of these measures. A special feature of Regulation 2078/92 is that, although the participation in agri-environmental programmes is voluntary for farmers, it is obligatory for the Member States to implement such programmes. It is thus the first common European framework for national policies in the agri-environmental field (Brouwer and Lowe 1998).

Regulation 2078/92 is a very diverse and broad instrument that should be sufficiently flexible to consider the differences in geographical conditions, agricultural production systems and rural traditions within the territory of the European Union. Because of these diverging regional circumstances, it is obvious that the elaboration and implementation of Regulation 2078/92 takes place on a national, regional or even local level. As a result, Regulation 2078/92 includes about 2200 distinct measures incorporated in 127 programmes. Programmes can be described as the way national or regional governments implement Regulation 2078/92, whereas measures are the specific agri-environmental actions introduced at a local level as components of national or regional programmes (see Biehl 1999). The European Commission has established a number of aid schemes that should be regarded by the Member States when applying for financial aid for these programmes. The aid schemes are described in Article 2.1 and 2.2 of the Regulation and they are shown in Table 1.

---

<sup>1</sup> Objective 1 regions are those whose development is lagging behind in the sense that their per capita GDP is less than 75% of the Community average over the past 3 years (EC 1995-2000).

**Table 1: Scheme Objectives Eligible for Aid under Regulation 2078/92.**  
 Source: CEC (1992), Buller (2000), Deblitz (1999)

| <b>Article 2.1 of Regulation 2078/92: Scheme objectives eligible for aid</b> |  |
|--|--|
| <b>a</b>   | To reduce substantially the use of fertilisers and/or plant protection products, or to maintain the reductions already made; or to introduce or to continue with organic farming.  |
| <b>b</b>   | To change, by means other than those referred to in (a), to more extensive forms of crop, including forage production; or to maintain extensive production methods introduced in the past; or to convert arable land into extensive grassland. |
| <b>c</b>   | To reduce the number of sheep and cattle per forage area.  |
| <b>d</b>   | To use other farming practices compatible with the requirements of protection of the environment and natural resources, as well as maintenance of the countryside and landscape; or to rear animals or local breeds in danger of extinction.   |
| <b>e</b>   | To ensure the upkeep of abandoned farmland or woodlands.   |
| <b>f</b>   | To set aside farmland for at least 20 years with a view to use it for purpose of the environment, in particular for the establishment of biotope reserves of natural parks or for the protection of hydrological systems.                      |
| <b>g</b>   | To manage land for public access and leisure activities.   |
| <b>Article 2.2 of Regulation 2078/92</b>                                     |  |
| Training and demonstration projects for farmers.                             |  |

The table makes clear that the aid schemes comprise indeed a wide range of agricultural practices promoting environmentally friendlier ways of farming. Next to this variety of agri-environmental measure, there are also different strategies how to implement them. Buller (2000) distinguishes between four broad models of implementation. Firstly, there are the targeted or zonal measures that aim at specific landscape types, natural regions or farming systems and at farmers located in a particular zone. Examples for this type of measure are the Environmentally Sensitive Area (ESA) schemes in Denmark and the United Kingdom. Targeted or zonal measures are applied in most of the Member States. Secondly, there are wide horizontal schemes that cover whole nations or regions addressing certain eligibility criteria such as, for example, grassland in the "Prime à l'herbe" in France. A third type of implementation strategy is a broad regulatory framework that generally consists of a basic initial payment to participating farmers and a number of additional aid-schemes requiring further restrictions and accordingly higher payments. The Irish Rural Environmental Protection Scheme (REPS) is an example of this type of implementation strategy. Fourthly, there are measures that focus on specific actions, such as the conversion and maintenance of organic farming, the protection of local breeds in danger of extinction, or training and demonstrations projects for farmers. In general, it can be observed that schemes that demand changes in agricultural techniques involve higher payments than those focussing on the maintenance of existing extensive practices.

The EC has also proposed a categorisation of the aid schemes into five groups: i) organic farming; ii) farming with environmental improvements; iii) maintenance of low intensity systems; iv) non-productive land management; v) training and demonstration projects. The preferences for these schemes among the Member States appear to vary significantly. For example, Mediterranean countries tend to use Regulation 2078/92 mainly for non-productive land management, which can be seen as a complementary source of income for farmers. Belgium, Denmark and Italy distinguish themselves from other Member States by allocating large proportions of

their 2078/92 budget to organic farming, whereas Sweden and The Netherlands are in favour of training and demonstration projects (Buller 2000).

By April 1997, 1.3 million contracts have been signed. This reflects around 18% of farms and 17% of total Utilisable Agricultural Area (UAA) in the EU (Buller 2000). The fact that the number of farms is slightly higher than the area under contract shows that there is a tendency of small-scale farmers to participate in the programmes.

It has to be mentioned that environmental concerns are not the only objective of Regulation 2078/92. Article 1 of the Regulation establishes three major goals: firstly, accompanying the changes to be introduced under the CAP reform in 1992; secondly, contributing to the Community's policy objectives regarding agriculture and the environment; and thirdly, contributing to providing an appropriate income for farmers (CEC, 1992). The first goal refers to the basic purposes of the 1992 CAP reform, namely the reduction of overproduction and its enhancing costs, the reduction of market support measures and the introduction of a system of direct payments. The second goal addresses the growing concern about the negative effects of agriculture on the environment such as water pollution, biodiversity loss and landscape change. It is furthermore the first effort to comply with the Maastricht Treaty that requires EU environmental policy to be integrated into all other EU policies. The third goal is concerned with the maintenance and protection of extensive farming practices, not only against intensification but also against agricultural decline and withdrawal (Buller 2000).

In the light of global liberalisation of agricultural trade, especially the last goal stated in Article 1 of Regulation 2078/92, is a critical factor, since it can be interpreted as justifying the continuation of funding and subsidisation of European agriculture disguised as 'green' CAP (Buller 2000). Nevertheless, agri-environmental support payments to farmers are accepted according to the GATT agreement on agriculture. The WTO member countries agreed upon a reduction of domestic support measures to agriculture by 20 per cent between 1995 and 2000 with respect to the support level in 1986-88. This reduction only refers to the so-called Amber Box support measures, which have the most disturbing effect on agricultural production and hence also on trade. A typical example of an Amber Box support measure is price support, which gives farmers direct economic incentives to expand or reduce their production. Agri-environmental policies belong to Green Box support measures. These measures are meant to have only a very small effect on production and trade, since the payments are supposed to be totally decoupled from production. Other examples of Green Box measures are e.g. general services, such as research or pest and disease control, domestic food aid, and compensation payments for natural disasters. There are also Blue Box support measures, which provide payments on the basis of a fixed amount of hectares or livestock in the frame of production limiting programmes (Silvis and Rijswick 1999).

There are still other criticisms about the current structure of agri-environmental policy in the EU. The most important one for policy makers is that, due to the wide variety of implementation strategies, it is rather difficult to carry out cross-national comparisons of scheme effectiveness and to evaluate the economic efficiency of the schemes in general. Another criticism is that the environmental policy target is in many cases far too broad and not adequately identified, so that potential positive effects on the environment cannot be evaluated. Furthermore, it is argued that Regulation 2078/92 is poorly integrated with other CAP policies. For instance, the maize premium of the regular CAP programme is in many cases higher than the grassland premium under the agri-environmental policy (Buller 2000).

Realising the critical points of agri-environmental policymaking, the EC has shifted the importance more and more towards the environmental objective, the second goal of Regulation 2078/92. In order to manifest these changes legally, the EC introduced in 1999 a new tool, viz. is the *Integrated Rural Development Regulation* (Regulation 1257/99). This new regulation integrates not only Regulation 2078/92, but also other rural measures such as the Less Favoured Area scheme. In this new regulation, income support to farmers is no longer mentioned, and environmental goals are clearly specified for farmers who want to participate in agri-environmental policy programmes (Lowe and Baldock 2000).

### **3. Environmental Indicators**

For a proper quantitative policy assessment, we have to resort to reliable indicators. The OECD (1997) defined three major functions of environmental indicators in agriculture. They should provide information to policy makers and the general public about the state of the environment influenced by agriculture. Furthermore, they have to help policy makers to better understand the cause-effect loops between agricultural activity and the environment. Finally, they have to assist in the evaluation of the effectiveness of agri-environmental policy instruments. In order to comply with these three demands, the OECD has proposed to apply a so-called Driving force-State-Response (DSR) framework. Driving forces are the factors that cause environmental conditions to change, such as input and output levels of farm production, agricultural land use, and also natural processes and meteorological conditions. The state describes the actual condition of the environment, like, for example, the nutrient level in ground and surface water or the number of protected species in a certain area. Response refers to the reactions of policy makers and groups of the society to the state of the environment.

Although the actual state of the environment would be the most appropriate indicator for policy evaluation, it is, especially in agriculture, also the most difficult one to assess. This has several reasons. One of the most important ones is the time and space dimension inherent to the cause-effect loop between agricultural production and the state of the environment. This means that the effects of agricultural pollution might become visible only after a number of years or that they spread out over long distances through, for example, water or air (Deblitz 1999). Another prominent reason is that the assessment of state indicators is in most cases rather costly.

The most appropriate alternative is to take the driving force indicator as a measure for the effectiveness of agri-environmental policy. In this case, the driving force indicators are agricultural practices that have a certain effect on environmental quality. The indicators used in this study are the same as in the so-called FAIR research project (see Section 4), since this project provides the data as input for the meta-analysis carried out in our study. The FAIR research project developed 12 different indicators based on 9 particular agricultural practices. The agricultural practices were selected according to three criteria, viz. relevance, reliability and realisability. Relevance implies the correspondence of the agricultural practices to the specific goal and actions of Regulation 2078/92. Reliability requires that the impact of the agricultural practices on the environment must be well known and scientifically proved. Finally, realisability refers to the availability of the appropriate data (Andersen *et al.* 1999).

For the purpose of meta-analysis a significant minimum amount of systematic and common data is needed. Since not all of the 12 indicators comply with this



requirement, we were forced to deploy only three of them, namely mineral nitrogen fertiliser, livestock density and grassland area per utilisable agricultural area. Our choice of indicators is hence solely based on data availability.

The actual relationship between the agricultural practices serving as our indicators and environmental quality is described in several scientific studies. Andersen *et al.* (1999) give a concise literature overview of these relationships. In the following, a short summary of this literature overview is given.

- *Mineral nitrogen fertiliser*: The excessive use of N-fertiliser can change the botanical composition of grassland by favouring particular species against others. This in turn harmfully influences specific bird populations that use grassland as their breeding and feeding habitat. Furthermore, intensive mineral N-fertilisation increases the nitrogen stock in the soil, which results into a rate of nitrification that is higher than the nitrogen demand of the current crop. As a consequence, the surplus of nitrogen will leach into groundwater. In order to comply with European standards for drinking water, the level of nitrate in groundwater must not exceed 50 mg/litre.

The mineral N-fertiliser indicator is measured in kg N-fertiliser per hectare and it has a negative relationship with the state of the environment. A decreasing value of the indicator is therefore preferable.

- *Livestock density*: A large amount of livestock per agricultural area is equivalent to high levels of the manure and slurry on this area. This, in turn, is directly related to leaching of nitrate into groundwater resources. However, the actual relationship between livestock density is found out to be bell-shaped. This means that livestock densities that are either too high or too low result in a degradation of the traditional ecological system. In our case, the second half of the bell-shaped curve is of importance, which implies that livestock densities have to be reduced in order to improve environmental quality.

The livestock density indicator is measured in total livestock units (LU) per hectare of utilisable agricultural area. It has a negative relationship with the state of the environment, which means that decreasing livestock density is (in general) favourable for the environment.

- *Grassland area per utilisable agricultural area (UAA)*: In comparison with arable land, grassland has many environmental advantages. First of all, the loss of nitrogen under grassland is significantly smaller than under arable land. Since ploughing accelerates the mobilisation of nitrate, it is favourable to prevent the conversion of grassland to arable land. Furthermore, the maintenance of extensive grassland is desirable because not only intensification but also abandonment negatively affects the variety of faunal and floristic species of grassland, which again has an unfavourable impact on grassland birds. Finally, grassland is an ideal measure for the prevention of soil erosion through wind and water. From all these facts it can be concluded that there is a direct positive relationship between the share of grassland and the state of the environment, which means that an increase in the indicator is beneficial to the environment. The grassland indicator is measured as percentage of grassland per UAA.

#### 4. Input Data: Case Studies of an EU Project

The case studies used in our empirical meta-analysis are the results of a three year project<sup>2</sup> funded by the European Union about the implementation and effectiveness of agri-environmental schemes established under Regulation 2078/92 (for the full project report see Schramek *et al.*, 1999). The project includes nine EU countries, namely Sweden, Denmark, Germany, Great Britain, France, Austria, Spain, Portugal and Greece. Additionally, it considers Switzerland for comparing experiences of non-EU-members that apply agri-environmental policies comparable to those of the EU. The research group was characterised by multidisciplinary and consisted of agricultural economists, general economists, ecologists, geographers, landscape planners and sociologists.

The data collection took place through farm surveys based on a uniform questionnaire. 22 case study areas were selected, two in each country, except for Sweden where four case study areas were selected. In total, 1000 farms were interviewed, 50 in each case study area (and 25 in the Swedish cases study areas). The study areas cover a wide range of European landscape types and different agri-environmental programs and are selected according to a limited number of agri-environmental issues, such as e.g. contamination of groundwater and soil, or biodiversity.

The objective of this research project was "*...to develop common and appropriately regionalised operational methodologies, and to apply these methodologies in order to analyse the implementation and effectiveness of EU-agri-environmental schemes established under Regulation 2078/92.*" (Schramek *et al.* 1999, 1). With the help of the questionnaire, the research group was not only able to identify and analyse farmers' participation in and attitudes towards agri-environmental policies, but they were also able to trace the environmental and socio-economic impacts of EU policies. For the purposes of this paper, we will mainly focus on the results of the environmental impact analysis.

The input data for the meta-analysis stem from the research group's analysis of the environmental effects of Regulation 2078/92 that was carried out on a case study level (for a detailed description of this analysis we refer to Andersen *et al.* 1999). As mentioned in the previous section, in the project 12 environmental indicators were proposed and developed on the basis of certain agricultural practices, of which three are selected for the analysis in the present paper. The reasons for the limited availability of data from the other 9 indicators lie in the fact that not all indicators are applicable to all case study areas and that response rates were too low in some areas to ensure reliable inferences. The indicators for the agricultural practices '*reduction in the use of mineral N-fertiliser (kg/ha)*', '*reduction of livestock density (LU/ha)*' and '*increase of grassland area with respect to total agricultural area (% grassland/UAA)*' are reflected by the average change rates per case study area of these practices over a five year period (1993-1997). The data for the calculation of these average change rates are taken from the individual farm questionnaires. The farmers interviewed are classified into two groups. On the one hand, there are farmers who are eligible for and participating in agri-environmental programmes, and on the other hand, there are farmers who are also eligible but not participating. The approach of comparing the behaviour of participating farmers to that of non-participating farmers makes it possible to directly identify the environmental impact of the programmes concerned. In the research process of the FAIR project the average change rates of the two groups

---

<sup>2</sup> Project FAIR 1 CT95-274

of farmers were compared statistically on a case study area level. The statistical test methods used for the comparison of the two groups are the t-Student test and the U-Mann Withney test.

The subdivision of the interviewees into participants and non-participants can be interpreted as a quasi-experimental research design. In this case, participating farmers act as the experimental group and non-participating farmers as the control group, be it that both are eligible. The non-random assignment of subjects, on the basis of self-selection, may cause some bias in the analysis, but since there was no other way of creating a data base, this shortcoming has to be accepted. It means that our results have to be interpreted with caution. The structure of experimental and control group is a proper base for conducting a meta-analysis, where so-called effect sizes are calculated, which reflect the relative difference between these two groups. Further explanations about meta-analysis as it is carried out in the present paper are offered in Section 5.

The comparisons between the selected indicators' average change rates of participants and non-participants on a case study level carried out by the FAIR research team gave the following results.

- *Kg N-fertiliser per hectare*: This indicator appears to be relevant for most of the 22 case study areas. However, due to lack of enough data only nine case study areas could be analysed, and these are the following: i) Great Britain-Cambrian Mountains; ii) Great Britain-Devon Countryside; iii) Germany-Rhoen; iv) Germany-Wetterau; v) Denmark-Viborg County; vi) Denmark-Vestsjaelland; vii) Spain-Sahagún; viii) Austria-Nordburgenland, ix) Greece-Larisa. Significant and expected results -meaning that the average change rates in using N-fertiliser of participants are negative and significantly higher than those of non-participants- are found in Spain-Sahagún and Greece-Larisa at a 5%-level and in Germany-Wetterau and Great Britain-Cambrian Mountains at a 10%-level. In all other case study areas the results are insignificant. An unexpected positive (but insignificant) average change rate of participating farmers is found in Denmark-Vestsjaelland.
- *Livestock Unit per hectare*: This indicator is relevant for 14 case study areas, of which two had to be dropped because of limited data availability and because of a significant change in the reference area. The following 12 case study areas are ultimately included in the analysis: i) Great Britain-Cambrian Mountains; ii) Great Britain-Devon Countryside; iii) Denmark-Viborg County; iv) Denmark-Vestsjaelland; v) Portugal-Moura; vi) Portugal-Castro Verde; vii) Austria-Nordburgenland; viii) Austria-Osttirol; ix) Switzerland-Schwarzwasser; x) Switzerland-Erlach/Seeland; xi) Germany-Rhoen; xii) France-Bocage-Avesnois. The only significant result -meaning that the average change rate of participants is negative and significantly higher than that of non-participants- is found in Germany-Rhoen at a 5%-level. In all other case study areas, the results are insignificant. Unexpected positive (but insignificant) average change rates of participating farmers are found in Switzerland-Schwarzwasser, Denmark-Vestsjaelland, Austria-Nordburgenland and Austria-Osttirol.
- *Grassland area (%) per UAA*: This indicator concerns 18 case study areas. However, only 13 case study areas could be analysed again due to limited data availability and significant changes in the reference area. The analysed case study area are the following: i) Great Britain-Cambrian Mountains; ii) Great Britain-

Devon Countryside; iii) Denmark-Viborg County; iv) Denmark-Vestsjaelland; v) Sweden-Enköping; vi) Sweden-Offerdal; vii) Sweden-Vallakra; viii) Austria-Nordburgenland; ix) Austria-Osttirol; x) Switzerland-Schwarzwasser; xi) Switzerland-Erlach/Seeland; xii) Germany-Rhoen; xiii) France-Bocage-Avesnois. The only significant result with an expected positive average change rate that is higher for participants than for non-participants is found in Sweden-Enköping at 10%-level. Another significant but paradoxical result, namely that the average change rate of participants is negative and higher than that of non-participants, is found in Great Britain-Devon Countryside. In all other case study areas the results are insignificant. In Switzerland-Schwarzwasser, France-Bocage-Avesnois, Great Britain-Cambrian Mountains, Sweden-Offerdal and Sweden-Vallakra the average change rates are unexpectedly negative (but insignificant).

Summarising, the number of expected and significant results of the change rates for the tree indicators is rather limited. At a 5%-level, two out of nine results are significant for N-fertiliser, one out of 12 result is significant for livestock density, and for grassland no significant results was found at all. However, it has to be kept in mind that the sample sizes in the individual case study areas tend to be rather small, which increases the probability of accepting the null-hypothesis although it may be false. This problem will be further elaborated in the following section.

Nine, 12 respectively 13 observations for the three indicators N-fertiliser, livestock density and grassland area are available for the analysis in this paper. More observations per indicator would certainly improve and strengthen our analysis because they would allow a more varied and differentiated investigation. However, since we are bound to the limited observations available, the analysis may be seen as a first exploration to apply the techniques of meta-analysis to agri-environmental policy evaluation. The statistical procedure of the meta-analysis employed in this paper is described in the following section.

## **5. Methodology of Research Synthesis**

### **5.1 Introduction**

The methodology adopted for our empirical case study is based on meta-analysis. Meta-analysis has already a quite remarkable history in psychology and medical science and found only recently its way to regional and environmental economics. The development of meta-analysis in psychology and medical science is for the main part due to large amounts of case studies on the same scientific issue performed in an experimental and largely standardised context, which forms a perfect base for statistically-based meta-analysis. The lack of experimental and standardised conditions in many fields of the social sciences (including economics) is actually the major criticism of applying meta-analysis to social science issues. In order to be able to compare existing research results in a strict statistical way, studies should involve quantitative factors and identical units, or at least results that can be transformed into some common unit or index (Van den Bergh and Button 1997).

Because of their quasi-experimental approach, the results from the case studies carried out in the FAIR project form, for environmental economic purposes, suitable inputs for meta-analytic research. At this point, the potential additional value of applying meta-analysis to these case studies has to be identified.

The previous section presented the results of the analysis of the average change rates in the individual case study areas of the FAIR project, which show in many

cases insignificant results. It was mentioned that this could be due to relatively small sample sizes. Standard statistical theory would tell us that parameter estimates from large sample sizes are more robust than those from small sample sizes, because the variance around parameter estimates from large samples is smaller (Shadish and Haddock 1994). Consequently, estimates from large samples tend to have higher values of significance. On the other hand, estimates obtained from rather small samples are, due to their larger variances, subject to the risk of Type II errors, which means accepting the null hypothesis, even though it may be false (Hunter and Schmidt 1990). This problem is aggravated, if the estimated population effect is small. Summarising case study results from small samples by simple vote-counting procedures, which means counting significant results only, might lead to the conclusion that the average effect of the intervention is not significantly different from zero (Hedges and Olkin 1985). Meta-analysis artificially increases the sample size by pooling, so to speak, all sample sizes from the individual studies for the calculation of the average effect size. An effect size is a generic term that refers to the magnitude of an effect or, more generally, the size of the relation between two variables (Cooper and Hedges, 1994). (A detailed description of effect sizes is given in Section 5.2). Meta-analysis does hence not take into account the significance level of the individual studies, but only their effect sizes. Because of the increased sample size, the calculated average effect size is generally significantly different from zero (Ijskes 1999). In our case, we will test if the average change rates of participating and non-participating farmers with respect to the three indicators are indeed significantly different from each other, although most of the original results are insignificant.

Another question that meta-analytic techniques are able to answer is whether individual studies share a common effect size, or, in other words, whether there is a single overall effect size that describes the general magnitude of the intervention. If this is not the case, meaning that the individual effect sizes are too heterogeneous to support the hypothesis of a common effect size, there must be factors at work that are responsible for the variations among the individual effect sizes. The identification of these factors is another task of the meta-analysis carried out in this paper.

### **5.2 The Effect Size**

Several different forms of effect sizes can be found in the current meta-analytical literature. In the medical, social and psychological sciences, the areas that are considered to be the traditional disciplines in the application of meta-analysis, two types of effect sizes are most commonly used: the d-type and the r-type effect sizes. The most famous effect sizes of the d-type are Hedges'  $g$ , Cohen's  $d$  and Glass's  $\Delta$ . An example of an r-type effect size is the correlation coefficient  $r$ . Effect sizes of the d-type are all standardised mean differences of control and experimental groups, which differ from each other with respect to the way of standardisation. Hedges'  $g$  uses the pooled (experimental *plus* control group) standard deviation that is calculated with degrees of freedom, which are the total number of observations minus the number of groups, for standardisation. Cohen's  $d$  standardises also by the pooled standard deviation but uses the total number of observations instead of degrees of freedom for the computation. Finally, Glass's  $\Delta$  uses only the standard deviation of the control group for standardisation (Rosenthal 1991, 1994).

The choice of which effect size to apply depends on the one hand on the type of data available but also on personal preferences. In our case, the original studies compare two groups. They also report the means, standard errors and sample sizes of these groups such that it is most appropriate to calculate an effect size of the d-type.

This analysis employs Hedges'  $g$  as its effect size. Effect sizes of the  $d$ -type and of the  $r$ -type are actually convertible to each other. The correlation coefficient,  $r$ , is hence just another way of interpreting the effect sizes of the  $d$ -type (Hedges and Olkin 1985). An interpretation of effect sizes of the  $d$ -type is the following.

An effect size of the  $d$ -type reflects the difference between an experimental and control group in such a way that it is independent of sample size and unit of measurement. In fact, the effect size gives the difference between an experimental and control group in standard deviation units (Rosenberg *et al.* 1997). Hedges and Olkin (1985) interpret the effect size as the  $z$ -score of the normal cumulative distribution function, where its respective  $\Phi(z)$ -value is the proportion of control group scores that is less than the average score of the experimental group. For example, an effect size of 0.3 signifies that the score of an average individual of the experimental group exceeds the score of 62% ( $\Phi(0.3) = 0.62$ ) of the individuals of the control group. Rosenberg *et al.* (1997) give Cohen's convenient rule of thumb about the interpretation of effect sizes. This rule says that 0.2 implies a small effect, 0.5 a medium effect and 0.8 a large effect. Everything above 1.0 is considered to be a very large effect.

### 5.3 Meta-analysis in Four Steps

The meta-analysis performed in this paper is divided into four steps. The first step is the calculation of effect sizes for each case study area with respect to the selected environmental indicators. The next step is the combination of these effect sizes for each environmental indicator. Thirdly, it has to be investigated whether the estimated effect sizes are homogeneous, which means whether the effect sizes from the individual case studies share a common effect size. This is done by testing the null-hypothesis that there is no variation among the effect sizes. If this test is rejected, the fourth step has to be carried out, and that is the moderator analysis. The description of the statistical procedure is based on Hedges and Olkin (1985), Rosenthal (1991, 1994) and Shadish and Haddock (1994).

#### 1. Step: Calculation of the Effect size

As mentioned above, this analysis employs Hedges'  $g$  as its effect size. Hedges'  $g$  is calculated according to the following formula:

$$g = \frac{M_E - M_C}{S_p}, \quad (5.1)$$

where  $M_E$  is the mean of experimental group and  $M_C$  the mean of the control group.  $S_p$  is the pooled sample standard deviation computed as

$$S_p = \sqrt{V_p} = \sqrt{\frac{(N_E - 1)V_E + (N_C - 1)V_C}{N_E + N_C - 2}}, \quad (5.2)$$

where  $V_E$  and  $V_C$  are the variances of the experimental and control group and  $N_E$  and  $N_C$  the experimental and control group sample sizes, respectively.

Rosenthal (1991, 1994) and Hedges and Olkin (1985) point out that  $g$  is negatively biased, especially when sample sizes are small and population effects are large. Because of the small sample argument, our analysis uses the adjusted, unbiased  $g$ , viz.  $g^u$ , that is obtained by applying:

$$g^u = g * c(m), \quad (5.3)$$

where

$$c(m) = 1 - \frac{3}{4(m) - 1}, \quad (5.4)$$

and  $m$  are the degrees of freedom computed from the experimental and control group ( $N_E + N_C - 2$ ). However, in our analysis the actual difference between  $g$  and  $g^u$  turns out to be rather small. In our calculations,  $c(m)$  lies around 0.98, which is close to one and hence almost negligible.

### *Step 2: Combining Effect Sizes*

It was already noted above that larger samples produce more significant and reliable estimates. It is hence suitable to weight the effect sizes of large sample studies more heavily before combining them. According to Shadish and Haddock (1994), the most appropriate weight is the inverse of the variance of the respective effect sizes, as shown in the following formula.

$$w_i = \frac{1}{v_i}, \quad (5.5)$$

where  $w_i$  is the weight and  $v_i$  the variance of the  $i$ -th effect size calculated according to formula (5.2).

The combination of the different Hedges'  $g$ 's obtained from  $k$  case studies gives the average effect size,  $\overline{G}_\bullet$ , that is calculated as

$$\overline{G}_\bullet = \frac{\sum_{i=1}^k w_i g_i}{\sum_{i=1}^k w_i}. \quad (5.6)$$

In order to find out whether the average effect size is significantly different from zero, or, in other words, whether there is a significantly positive or negative average effect due to the intervention, the confidence interval of the average effect size has to be calculated. This requires the calculation of the average effect size's standard error,  $s_\bullet$ , that is given by

$$s_\bullet = \sqrt{v_\bullet} = \sqrt{\frac{1}{\sum_{i=1}^k w_i}}, \quad (5.7)$$

where  $v_\bullet$  is the average effect size's variance. Subsequently, the confidence interval can be computed according to

$$\overline{G}_{\bullet} \pm C_{\alpha} * s_{\bullet}, \quad (5.8)$$

where  $C_{\alpha}$  is the critical value of the standard normal distribution.

Alternatively, the null hypothesis that there is no average effect can be tested with the Z-statistic, and that is

$$Z = \frac{\overline{G}_{\bullet}}{s_{\bullet}}. \quad (5.9)$$

If Z exceeds 1.96, the 95 percent two-tailed critical value of the standard normal distribution, the null hypothesis can be rejected and it can be concluded that the intervention has a significant average effect.

### *Step 3: Test on Homogeneity of Effect Sizes*

Equation (5.6) assumes that all individual studies share a common effect size. This is certainly a very strong assumption and in most cases this is actually not the case. The test on the effect sizes of all individual studies being indeed not homogeneous is called the Q-test and it is represented by the following formula:

$$Q = \sum_{i=1}^k \frac{(g_i - \overline{G}_{\bullet})^2}{v_i} \quad (5.10)$$

If the value of Q exceeds the upper tail critical value of the  $\chi^2$ -square distribution with k-1 degrees of freedom, it has to be assumed that the effect sizes of the individual studies are not homogeneous and that the individual studies do not share a common effect size.  $\overline{G}_{\bullet}$ , as calculated in equation (5.6), has therefore to be interpreted as the mean of the observed effect sizes and not as a single effect parameter.

The heterogeneity of the effect sizes of the individual studies shows that there must be factors influencing the magnitude of the effect sizes. These factors are called moderator variables. The analysis of moderator variables is described in the next step.

### *Step 4: Analysis of Moderator Variables*

Moderator variables are the factors that determine the variations in the effect sizes among the individual studies. Another interpretation of moderator variables is that they are able to identify important study characteristics. In our case, moderators should explain the variations of the policy effect in the different case study areas. In other words, they should reflect the reasons why in some case study areas there is a larger difference in behaviour between participating and non-participating farmers with regard to a particular indicator than in other case study areas.

In general, moderator variables can roughly be categorised into three groups. Firstly, there are moderators based on the underlying theoretical framework. In our case, an example for a moderator of the first type would be the premium level. Theoretically, it can be assumed that higher premium levels would induce larger changes in behaviour with respect to the particular agricultural practice indicators. Secondly, there is the group of moderators including variables that reflect the setting of the particular case study. Variables that reflect the setting of a case study describe



country or time specific characteristics. Thirdly, there is the group of moderators that refer to methodological issues of the primary case studies. These variables represent the way in which the analysis in the primary study is carried out. Examples are the statistical method used, the functional form chosen, or the type of data employed in the primary study. In the present paper, the individual case studies will all apply the same statistical technique. This means that methodological moderators are not supposed to be very important in our case.

Certainly, the list of potential moderator variables is very long and again, the availability of information is the determining factor of which moderator variables to choose. The analysis in this current paper tests the existence and importance of following moderator variables.

*I) Average premium per hectare:* Theoretically, higher premiums would imply that farmers are more stimulated to change their behaviour with respect to the relevant agricultural practice indicators. Therefore, higher premiums would be related to larger effect sizes. The FAIR project reports average premiums per farm and average farm sizes of participating farmers for all case study areas. The moderator variable *average premium per hectare* is calculated by dividing average premium per farm by average farm size of participating farmers for all relevant case study areas. With this moderator it is tested whether higher premiums do indeed result into higher effect sizes.

*II) Average farm size:* Larger farms are supposed to be more innovative and more creative with respect to attaining alternative sources of income. Large farms have more hectares at their disposal. This means that they have a higher probability that some of their land is being located in eligible area or that some of their land is marginal anyway. (Total agri-environmental payments per farm will be higher for large farms, meaning that it is more attractive for them to sign a contract). With this moderator we want to test if the variable farm size has an effect of the magnitude of the effect size.

*III) Absolute level of indicator in 1997:* Case study areas that have in general a relative low (for N-fertiliser and livestock density) respectively high (for grassland) level of the indicator might have lower change rates of participating farmers and hence lower effect sizes. (The level of the indicator for the starting year (1993) would hence be more suitable. Unfortunately, not reported in the FAIR study.)

In the FAIR project, all case study areas are categorised into four groups, each of them describing the characteristics of the agricultural production structure in that area. The four different categories are intensive arable farming, extensive arable farming, intensive husbandry farming and extensive husbandry farming. Unfortunately, the number of observations available to us is not large enough for using this differentiated categorisation in one moderator analysis. Therefore, we had to simplify this categorisation into the moderators intensive versus extensive farming and arable versus husbandry farming and perform two separate analyses on these two moderator variables. The moderators intensive versus extensive farming and arable versus husbandry farming are only tested for the indicator Nitrogen-fertiliser.

IV) *Intensive versus extensive farming*: With this moderator it is tested whether effect sizes in areas of intensive farming differ significantly from those in areas of extensive farming.

V) *Arable versus husbandry farming*: With this moderator it is tested whether effect sizes in areas of arable farming differ significantly from those in areas of husbandry farming.

The most basic way of performing a moderator analysis is as follows. First of all, the sample of effect sizes has to be subdivided into two (or more, depending on the number of observations) groups that are associated with a particular characteristic reflected by a moderator variable. Subsequently, a meta-analysis as described in Step 1 through 3 has to be performed on the separate groups. Additionally, two more Q-tests have to be carried out. Firstly, there is the Q-test on heterogeneity between the groups, the Q-between test. Secondly, there is the Q-test on heterogeneity within the groups, the Q-within test. The Q-between statistic tests the null hypothesis that there is no variation across the group mean effect sizes. In other words, it tests whether a particular moderator variable has indeed a significant influence on the effect size. The Q-between statistic given by the following formula:

$$Q_{\text{between}} = \sum_{i=1}^p \frac{(\bar{g}_{i\bullet} - \bar{G}_{\bullet})^2}{v_{i\bullet}}, \quad (5.11)$$

where  $p$  is the number of groups,  $\bar{g}_{i\bullet}$  the average effect size of the  $i$ th group,  $\bar{G}_{\bullet}$  the overall average effect size (formula (5.6), also called the grand weighted mean), and  $v_{i\bullet}$  the variance of  $\bar{g}_{i\bullet}$ , calculated according to formula (5.7), taking into account the observations in that particular group only.

The Q-within statistic is presented by the following formula:

$$Q_{\text{within}} = \sum_{i=1}^p \sum_{j=1}^m \frac{(g_{ij} - \bar{g}_{i\bullet})^2}{v_{ij}}, \quad (5.12)$$

where  $m$  is the number of observations in  $i$ th group,  $g_{ij}$  the  $j$ th effect size in the  $i$ th group, and  $v_{ij}$  its variance, according to formula (5.7), taking into account the observations in that particular group only. In fact, the Q-within statistic is the sum of the Q-tests (formula (5.10)) applied to every single group:

$$Q_{\text{within}} = Q_{w1} + Q_{w2} + \dots + Q_{wp}. \quad (5.13)$$

The sum of the Q-between and the Q-within statistic results in the overall Q-test applied to all observation (formula (5.10)):

$$Q = Q_{\text{within}} + Q_{\text{between}}. \quad (5.14)$$

In an ideal case, the selected moderator variable explains total heterogeneity such that most of the heterogeneity is between groups. If there is still heterogeneity within groups, the selected moderator variable is not able to explain all the variation among

the effect sizes. If the number of observations within the groups is still large enough, a moderator analysis can be performed within the groups. This procedure could continue until there is no within-group heterogeneity anymore.

## 6. Results of the Effect Size Analysis

This section presents the results of the meta-analysis applied to the evaluation of the three agri-environmental indicators N-fertiliser, livestock density and grassland area. Section 6.1 describes the outcomes of Step 2 and 3. Section 6.2 gives the results of the moderator analyses.

### 6.1 Combined Effect Sizes and Homogeneity Test

The outcomes of Step 2 (combining effect sizes) and Step 3 (test on homogeneity) as described in the previous section are reported in Table 2.

**Table 2: Results of Step 2 and Step 3**

|                     | k  | N<br>(N <sub>E</sub> , N <sub>C</sub> ) | Hedges<br>g | Var   | SE    | conf. interval |        | Q      | P <sub>(Q)</sub> |
|---------------------|----|---|-------------|-------|-------|----------------|--------|--------|------------------|
|                     |    |   |             |       |       | min            | max    |        |                  |
| <b>N-fertiliser</b> | 9  | 349<br>(242/107)                        | -1.573      | 0.236 | 0.154 | -1.874         | -1.272 | 52.24  | 0.000            |
| <b>Livestock</b>    | 13 | 630<br>(445/185)                        | -0.816      | 0.012 | 0.111 | -1.033         | -0.598 | 161.81 | 0.000            |
| <b>Grassland</b>    | 13 | 569                                     | -0.831      | 0.015 | 0.122 | -1.07          | -0.591 | 169.84 | 0.000            |

(k: number of case study areas; N: number of individual farmers; N<sub>E</sub>: number of individual farmers in experimental group (participants); N<sub>C</sub>: number of individual farmers in control group (non-participants); VAR: variance of Hedges' g; SE: standard error of Hedges' g).

The table shows that the combined effect sizes of all three indicators are significantly different from zero (the confidence intervals do not include zero). Here, the effect of the increased sample sizes becomes visible. Although most of the original case studies show insignificant results, the combined effect sizes demonstrate that there is an overall difference between the change rates of participating and non-participating farmers.

The effect sizes of the indicators N-fertiliser and livestock density have the expected negative sign. However, the sign of the effect size of the indicator grassland is unexpectedly negative. This result is paradoxical because the policy is meant to increase the area of grassland. The fact that the confidence interval does not include zero makes this result even more contradictory.

The indicator N-fertiliser has the highest average effect size, namely -1.57. According to the interpretation of effect sizes described in Section 5.2, 94% ( $\Phi(1.57) = 0.94$ ) of the change rates of non-participating farmers are lower than the average change rate of participating farmers. Applying Cohen's rule of thumb (see Section 5.2), it can be stated that an effect size of -1.57 reflects a very large effect of the policy intervention regarding the use of fertiliser. It has to be noted that the effect sizes do not say anything about the difference in the actual size of the change rates of participating and non-participating farmers but only about the percentage value at which the change rates of non-participants lie under the average change rate of participants.

The effect size for the indicator livestock density is -0.82. This means that 79% of the change rates of non-participating farmers are lower than the average change rate

of participating farmer. According to Cohen's rule of thumb, this effect size exhibits a large effect of the policy intervention with regard to livestock density.

However, the Q-test on homogeneity signifies at a very high significance level for all three indicators that the effect sizes of the individual case study areas are heterogeneous. This means that the case study areas do not share a common effect size, but that the calculated effect size is only the mean of the effect sizes in the individual case study areas.

## 6.2 Moderator Analyses

Since the calculated effect sizes do not pass the Q-test on homogeneity, a moderator analysis as described in Step 4 has to be carried out. Firstly, the moderator 'average premium per hectare', secondly, the moderator 'average farm size of participating farmers' and thirdly, the moderator 'average absolute value in 1997' (of the indicator) will be tested. Finally, the moderators 'intensive versus extensive farming' and 'arable versus husbandry farming' will be considered, but only for the indicator N-fertiliser.

### 1) Average Premium per Hectare:

The results of the moderator analysis 'average premium per farm' are shown in the following table.

**Table 3: Results of Moderator Analysis 'Average Premium per Hectare'.**

| 2 groups  | N-FERTILISER |       |                  | LIVESTOCK DENSITY |        |                  | GRASSLAND |        |                  |
|-----------|--------------|-------|------------------|-------------------|--------|------------------|-----------|--------|------------------|
|           | Hedges' g    | Q     | P <sub>(Q)</sub> | Hedges' g         | Q      | P <sub>(Q)</sub> | Hedges' g | Q      | P <sub>(Q)</sub> |
| < 40 ECU  | -1.31        | 22.17 | 0.00             | -0.51             | 106.17 | 0.00             | -0.81     | 66.06  | 0.00             |
| > 40 ECU  | -1.83        | 27.19 | 0.00             | -1.78             | 31.83  | 0.00             | -0.85     | 103.76 | 0.00             |
| Q between |              | 2.88  | <b>0.09</b>      |                   | 23.81  | 0.00             |           | 0.02   | <b>0.88</b>      |
| Q within  |              | 49.36 | 0.00             |                   | 138    | 0.00             |           | 169.82 | 0.00             |
| 3 groups  |              |       |                  |                   |        |                  |           |        |                  |
| < 30 ECU  | -0.80        | 8.24  | 0.02             | -0.41             | 38.47  | 0.00             | -0.64     | 63.24  | 0.00             |
| > 30 ECU  | -2.54        | 0.45  | <b>0.8</b>       | -0.60             | 69.15  | 0.00             | -1.46     | 9.229  | 0.03             |
| > 100 ECU | -1.23        | 18.13 | 0.00             | -1.92             | 24.35  | 0.00             | -0.26     | 81.45  | 0.00             |
| Q between |              | 25.42 | 0.00             |                   | 29.84  | 0.00             |           | 15.93  | 0.00             |
| Q within  |              | 26.82 | 0.00             |                   | 131.97 | 0.00             |           | 153.91 | 0.00             |

For the moderator 'average premium per farm' two kinds of analyses were carried out. In the first analysis, the effect sizes are divided into two groups. The first group comprises all case study areas where the average premium is less than 40 ECU per hectare, and the second groups includes all case study areas where average premium is larger than 40 ECU per hectare. For the indicators N-fertiliser and livestock density, the results are as expected, namely that higher average premiums per hectare result into higher effect sizes. (Remind that a higher effect size does not indicate higher actual change rates of the indicators for participating farmers, but only that a higher percentage of the change rates of non-participating farmers is lower than the average change rate of participating farmers). The Q-between test is highly significant for livestock density, and significant at a 10%-level for N-fertiliser. This means that the effect sizes of the two groups are significantly different from each other. However, the Q-within statistics still indicate heterogeneity among the effect sizes in the two groups. For the indicator grassland, the effect sizes of the two groups are not

significantly different from each other (the Q-between test cannot reject the null hypothesis of homogeneity).

Since the Q-within tests in the 2-groups analysis still indicates heterogeneity among effect sizes, a second analysis was carried. In this second analysis, we tried whether a division into 3 groups might improve the Q-within tests. The first group includes all case study areas where the average premium is less than 30 ECU per hectare, the second group contains all case study areas with an average premium between 30 ECU and 100 ECU per hectare, and the third group comprises all case study areas where the average premium per farm is above 100 ECU per hectare. As it is shown in the table, only for the indicator livestock density increasing premiums per hectare result into higher effect sizes. The Q-between test also rejects the null hypothesis of homogeneity among the average effect sizes of the three different groups. The Q-within statistic slightly decreased, but is still indicating heterogeneity of the effect sizes within the groups. For the indicator N-fertiliser, the second group shows the largest effect size, and it is additionally one of the few cases where the Q-within test indicates homogeneity for that group. For the indicator grassland, the Q-between test now signifies heterogeneity among the average effect sizes between groups. However, the unexpected negative effect sizes remain in all the groups.

Summarising, in the second analysis the Q-between tests indicate heterogeneity, which means that the moderator ‘average premium per hectare’ has a significant influence on the magnitude of the effect size. However, in the ideal case, additional to between-group heterogeneity, the Q-within tests should indicate homogeneity. This does not occur in this first moderator analysis. Unfortunately, the number of observations is not large enough for a more differentiated analysis.

## II) Average Farm Size of Participating Farmers

The results of the moderator analysis ‘average farm size’ are presented in the following table.

**Table 4: Results of Moderator Analysis ‘Average Farm Size’.**

| 2 groups  | N-FERTILISER |       |                  | LIVESTOCK DENSITY |        |                  | GRASSLAND |        |                  |
|-----------|--------------|-------|------------------|-------------------|--------|------------------|-----------|--------|------------------|
|           | Hedges' g    | Q     | P <sub>(Q)</sub> | Hedges' g         | Q      | P <sub>(Q)</sub> | Hedges' g | Q      | P <sub>(Q)</sub> |
| < 80 ha   | -1.54        | 21.85 | 0.00             | -0.92             | 100.85 | 0.00             | -0.87     | 103.71 | 0.00             |
| > 80 ha   | -1.59        | 30.36 | 0.00             | -0.72             | 60.12  | 0.00             | -0.81     | 66.07  | 0.00             |
| Q between |              | 0.03  | <b>0.87</b>      |                   | 0.84   | <b>0.36</b>      |           | 0.063  | <b>0.80</b>      |
| Q within  |              | 52.21 | 0.00             |                   | 160.97 | 0.00             |           | 169.78 |                  |
| 3 groups  |              |       |                  |                   |        |                  |           |        |                  |
| < 40 ha   | -1.23        | 18.13 | 0.00             | -1.92             | 24.35  | 0.00             | -0.26     | 81.45  | 0.00             |
| > 40 ha   | -1.55        | 4.66  | <b>0.1</b>       | -0.16             | 45.89  | 0.00             | -1.20     | 46.65  | 0.00             |
| > 100 ha  | -1.84        | 26.97 | 0.00             | -0.89             | 54.47  | 0.00             | -0.86     | 32.73  | 0.00             |
| Q between |              | 2.471 | <b>0.29</b>      |                   | 37.10  | 0.00             |           | 9.01   | 0.01             |
| Q within  |              | 49.77 | 0.00             |                   | 124.71 | 0.00             |           | 160.84 | 0.00             |

As in the previous case, we performed two kinds of analyses, one with two groups and another one with three groups. In the first analysis, the first group contains all case study areas where the average farm size of participating farmers is lower than 80 ha and the second group all case study areas where the average farm size of participating farmers is higher than 80 hectare. In this first analysis, the Q-between tests of all three indicators signify homogeneity between the effect sizes of the two

groups. This means that this first analysis does not support the assumption that the moderator variable ‘average farm size of participating farmers’ has a significant influence on the magnitude of the effect size.

In the second analysis, the first group contains those case study areas where the average farm size of participating farmers is lower than 40 ha, the second group those where average farm size is between 40 and 100 ha, and the third group the ones with an average farm size of higher than 100 ha. For the indicator N-fertiliser, the Q-between test still shows homogeneity of the average effect sizes of the three groups, indicating that even in this more differentiated analysis, average farm size of participating farmers do not seem to be influential for the magnitude of the effect size of this indicator. For the other two indicators, the Q-between test shows heterogeneity between the average effect sizes of the three different groups. However, the Q-within test still indicates in all cases heterogeneity among the effect sizes inside the groups. Unfortunately, the number of observations is not large enough for a more differentiated analysis.

### III) Average Absolute Value 1997

In this third moderator analysis, we divided the effect sizes of the different case study areas into two groups. For the indicator N-fertiliser, the first group contains those case study areas where the average absolute value in 1997 is lower than 40 kg/ha and the second group those where it is higher than 40 kg/ha. For the indicator livestock density, the first group comprises all case study areas with less than 1.5 Livestock Units per hectare on average in 1997 and the second group the ones with more than 1.5 Livestock Units per hectare. For the indicator grassland, the two groups are characterised by less than, respectively more than 50% grassland area per UAA in 1997. The results of the moderator analysis ‘average absolute value in 1997’ are shown in the following table.

**Table 5: Results of Moderator Analysis ‘Absolute Value 1997’.**

|           | N-FERTILISER |       |                  | LIVESTOCK DENSITY |           |        |                  | GRASSLAND |           |        |                  |
|-----------|--------------|-------|------------------|-------------------|-----------|--------|------------------|-----------|-----------|--------|------------------|
|           | Hedges 'g    | Q     | P <sub>(Q)</sub> |                   | Hedges 'g | Q      | P <sub>(Q)</sub> |           | Hedges 'g | Q      | P <sub>(Q)</sub> |
| <40 kg/ha | -1.11        | 22.85 | 0.00             | <1.5 LU/ha        | -0.56     | 81.31  | 0.00             | <50%      | -0.73     | 75.73  | 0.00             |
| >40 kg/ha | -1.93        | 22.28 | 0.00             | >1.5 LU/ha        | -1.10     | 74.64  | 0.00             | >50%      | -0.92     | 93.48  | 0.00             |
| Q between |              | 7.12  | 0.01             |                   |           | 5.86   | 0.02             |           |           | 0.63   | <b>0.43</b>      |
| Q within  |              | 45.12 | 0.00             |                   |           | 155.95 | 0.00             |           |           | 169.21 | 0.00             |

The Q-between test signifies for the indicators N-fertiliser and livestock density heterogeneity between the average effect sizes of the two different groups. This implies that the average absolute value in 1997 seem to have a significant influence on the magnitude of the average effect size. As it was supposed, the case study areas with a higher absolute level of the indicator have a higher average effect size than those with a lower level of that indicator. This means that in areas with a higher absolute value of the indicator in 1997, a higher percentage of the change rates of non-participating farmers lie under the average change rate of participating farmers. For the indicator grassland, the Q-between test reports homogeneity between the average effect sizes of the two groups. The Q-within tests show in all cases

heterogeneity among the effect sizes. However, again, the number of observations is not large enough for a more differentiated analysis.

*IV and V) Intensive versus Extensive Farming, Husbandry versus Arable Farming*

The results of the last two moderator analyses are given in table 6.

**Table 6:** Results of Moderator Analysis ‘Intensive-Extensive’ and ‘Arable-Husbandry’.

| N - F E R T I L I S E R |          |        |                  |                  |          |        |                  |
|-------------------------|----------|--------|------------------|------------------|----------|--------|------------------|
|                         | Hedges g | Q      | P <sub>(0)</sub> |                  | Hedges g | Q      | P <sub>(0)</sub> |
| <b>Intensive</b>        | -1.488   | 11.904 | 0.008            | <b>Arable</b>    | -1.477   | 43.349 | 0.000            |
| <b>Extensive</b>        | -1.666   | 40.000 | 0.000            | <b>Husbandry</b> | -1.871   | 7.681  | 0.006            |
| Q between               |          | 0.336  | <b>0.562</b>     | Q between        |          | 1.208  | <b>0.272</b>     |
| Q within                |          | 51.904 | 0.000            | Q within         |          | 51.032 | 0.000            |

The table shows that the effect size for intensive farming is slightly lower than that of extensive farming and that the effect size for arable farming is lower than that of husbandry farming. However, the Q-between test signifies that the null hypothesis of between-group homogeneity cannot be rejected in both cases. This means that according to this analysis, the fact that the case study area is characterised by intensive or extensive, respectively arable or husbandry, agricultural production structure does not have any influence on the magnitude of the effect size. The Q-within tests indicate, as in most of the previous moderator analyses, still heterogeneity among the effect sizes in the two groups.

## 7. Conclusions and Discussion

In this paper, a first effort was made to apply the statistical methods of meta-analysis to the question of agri-environmental policy evaluation in the European Union. Because of the limited data availability, this study can rather be seen as exploratory and as a first test of how meta-analytical techniques handle the data. In spite of that some general conclusions can be drawn on the basis of this analysis. Firstly, the fact that meta-analysis artificially increases the sample size becomes visible in the results of Step 2, the combination of the effect size. This means that, although most of the original case studies show insignificant differences between the change rates of participating and non-participating farmers, the combined effect sizes demonstrate that there is an overall difference between the change rates. In other words, there is an indication that the agri-environmental policy intervention has indeed a positive effect on the behaviour of participating farmers with respect to the chosen indicators. Furthermore, from the moderator analysis, it can be concluded that the variables ‘average premium per hectare’ and ‘average absolute value in 1997’ have a significant effect on the magnitude of the effect sizes, meaning that they influence the percentage level at which the change rate of non-participating farmers lie below the average change rate of participating farmers. In general, the effect sizes of the indicator N-fertiliser show the highest value. This could be explained by the fact that the reduction of N-fertiliser is easier to organise and less dependent on other conditions than the reduction of livestock density or the increase in grassland area.

The number of livestock kept by a farmer is rather susceptible to current prices of meat and livestock, which might outstrip the payments for agri-environmental programmes. The effect sizes of the indicator share of grassland area per UAA all show unexpected negative signs. This paradoxical results may be due to the fact that the indicator grassland area is a very broad measure, being subject to multiple decision making processes, also outside the agricultural sector as for instance in urban and landscape planning.

A prevailing problem throughout all moderator analyses is that the Q-within tests signify heterogeneity of the effect sizes with the different groups. The occurrence of this problem does actually underline the diversity of the European landscape and the differences in the structure of the agricultural sector, which is often emphasised by researchers trying to evaluate European agri-environmental policy. The methods of meta-analysis might be able to shed more light on this diversity if a larger enough number of observations would be available. It would then be possible to apply more advanced methods of meta-analysis, such as multi-factor analysis by taking into account two or more moderator variables or meta-regression analysis. With the number of observations available to us in this paper number, it is not possible to get any more sophisticated conclusions out of it. In their study they made for the European Commission, the researchers of the FAIR project give the advice to introduce monitoring programmes with which the behaviour of participants and non-participants can be compared. With such a quasi-experimental impact assessment, it should be easier to compare policy outcomes with policy objectives. Quasi-experimental case study results would also increase the number of potential input data for meta-analysis. Retrieving case studies using such an approach is certainly the most important task for improving and strengthening the meta-analysis as it is performed in this paper.

## References

- Andersen, E., J. Primdahl, J. Oñate, J. Aguirre, C. Cummings, B. Peco, J. Schramek, B. Klingler and K. Nickel (1999) Analysing the Environmental Effects of Regulation 2078/92. In: D. Biehl, H. Buller and G. Wilson (eds) *Implementation and Effectiveness of Agri-Environmental Schemes Established under Regulation 2078/92*. Project FAIR 1 CT95-274, Final Consolidated Report, Volume II (annexes), Institut fuer laendliche Strukturforschung (Research Institute for Rural Development), Frankfurt.
- Baaijens, S., and P. Nijkamp, Meta-analytic Methods for Comparative and Exploratory Policy Research, *Journal of Policy Modeling*, vol. 22, no. 7, pp. 821-858
- Bergh, J.C.J.M. van den, and K.J. Button (1997) Meta-analysis of Environmental Issues in Regional, Urban and Transport Economics, *Urban Studies*, Vol. 34, Nos 5-6, pp. 927-944.
- Beinat, E., and P. Nijkamp (1998) Environmental Rehabilitation: Efficiency and Effectiveness in Soil Remediation, *Studies in Regional and Urban Planning*, vol. 6, pp. 83-101.
- Beinat, E., and P. Nijkamp (eds.) (1999) *Multicriteria Analysis for Land-Use Management*, Kluwer, Dordrecht.
- Bergh, J.C.J.M. van den (1996) *Ecological Economics and Sustainable Development*, Edward Elgar, Cheltenham, UK.
- Bergh, J.C.J.M. van den, K. Button, P. Nijkamp and G. Pepping (1997) *Meta-analysis in Environmental Economics*. Kluwer, Dordrecht.
- Biehl, D. (1999) Agri-Environmental Policies in a Wider Theoretical and Institutional Context. In: Schramek, J., D. Biehl, H. Buller and G. Wilson (eds) *Implementation and Effectiveness of Agri-*



*Environmental Schemes Established under Regulation 2078/92*. Project FAIR 1 CT95-274, Final Consolidated Report, Volume I (main report), Institut fuer laendliche Strukturforschung (Research Institute for Rural Development), Frankfurt.

Brouwer, F.M. and S. van Berkum (1996) *CAP and the Environment in the European Union*. LEI-DLO, Wageningen Pers, Wageningen.

Brouwer, F.M. and P. Lowe (1998) CAP Reform and the Environment. In: Brouwer, F.M. and P. Lowe (eds.) *CAP and the Rural Environment in Transition*. Wageningen Pers, Wageningen.

Buller, H. (2000) The Agri-Environmental Measure 2078/92. In: Brouwer, F. and P. Lowe (eds) *CAP Regimes and the European Countryside: Prospects for Integration between Agriculture, Regional and Environmental Policies*, Oxford University Press, Oxford, pp. 199-220

CEC (1992), *Council Regulation 2078/92 on the Introduction and Maintenance of Agricultural Production Methods Compatible with the Requirements of the Preservation of the Environment and the Management of the Countryside*. Office for Official Publications of the European Community, Luxemburg.

Chomitz, K.M., and D.A. Gray (1996) Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize, *World Bank Economic Review*, vol. 10, no. 3, pp. 487-512.

Cooper, H. and L.V. Hedges (1994) *The Handbook of Research Synthesis*, Russel Sage Foundation, New York.

Deblitz, C. (1999) *Vergleichende Analyse der Ausgestaltung und Inanspruchnahme der Umweltprogramme zur Umsetzung der VO(EWG) 2078/92 in ausgewählten Mitgliedstaaten der EU*. Landbauforschung (agricultural research institute)Völkenrode, Sonderheft (special issues) 195, Braunschweig.

European Communities (1995-2000)  
[http://www.europa.eu.int/comm/regional\\_policy/activity/erdf/er1b1\\_en.htm](http://www.europa.eu.int/comm/regional_policy/activity/erdf/er1b1_en.htm) (2000-03-09).

Finco, A., and P. Nijkamp (2000) Sustainable Future of Cities, Future Land Use in Europe (J. Stilwell and H. Scholten, eds.), Kluwer, Dordrecht, 2000 (forthcoming).

Giaoutzi, M., and P. Nijkamp (1994) *Decision Support Models for Regional Sustainable Development*, Ashgate, Aldershot, UK.

Glass, G., B. MacGraw and M. Smith (1984) *Meta-analysis in Social Research*, Sage, Beverly Hill.

Groeneveld, R.A., and E.C. van Ierland (2000) *Economic Modelling Approaches to Land Use and Cover Change*. Report no. 410 200 045, Department of Economics and Management, Wageningen University, Wageningen.

Hedges, L.V. and I. Olkin (1985) *Statistical Methods for Meta-Analysis*. Academic Press, New York.

Hunter, J.E. and F.L.Schmidt (1990) *Methods of Meta-Analysis*. Sage Publications, Newbury Park.

Ijskes, E. (1999) A Meta-analytic Examination of the Antecedents of Budgetary Slack. Paper in progress (20<sup>th</sup> December 1999), Free University Amsterdam, Department of Accountancy.

Lier, H.N. van, C.F. Jaarsma, C.R. Jurgens, and A.J. de Buck (1994) *Sustainable Land Use Planning*. Elsevier, Amsterdam.

Lowe, P. and D. Baldock (2000) Integration of Environmental Objectives into Agricultural Policymaking. In: Brouwer, F. and Lowe, P. (eds) *CAP Regime and the European Countryside: Prospects for Integration between Agriculture, Regional and Environmental Policies*. (forthcoming)

Matarazzo, B., and P. Nijkamp (1997) Methodological Complexity in the Use of Meta-analysis for

- Empirical Environmental Case Studies. *International Journal of Social Economics*, vol. 34, no. 719, pp. 799-811.
- Miller, J.M. and A.J. Plantinga (1999) Modeling Land Use with Aggregate Data. *American Journal of Agricultural Economics*, 81, pp. 180-194.
- Nijkamp, P. (2000) Critical Success Factors for Soil Remediation Policy. *Journal for Environmental Policy and Law*, no. 1, pp. 81-98
- Nijkamp, P., and R. Vreeker (2000) Sustainability Assessment of Development Scenarios: Methodology and Application to Thailand. *Ecological Economics*, vol. 33, no. 1, pp. 7-27.
- OECD (1997) *Environmental Indicators for Agriculture*. OECD, Paris.
- Parks, P.J. and J.P. Schorr (1997) Sustaining Open Space Benefits in the Northeast: An Evaluation of the Conservation Program. *Journal of Environmental Economics and Management*, 32, pp. 85-94.
- Petitti, D.B. (1994) *Meta-analysis, Decision Analysis and Cost-Effective Analysis for Quantitative Synthesis in Medicine*. Oxford University Press, New York.
- Rosenberg, M.S., D.C. Adams, and J. Gurevitch (1997) *MetaWin* (statistical software for meta-analysis with resampling test). Sinauer Associates, Sunderland, Massachusetts.
- Rosenthal, R. (1991) *Meta-Analytical Procedures for Social Research*. Sage Publications, Newbury Park.
- Rosenthal, R. (1994) Parametric Measures of Effect Sizes. In : Cooper, H. and Hedges, L.V. (eds) *The Handbook of Research Synthesis*. Russel Sage Foundation, New York.
- Schramek, J. D. Biehl, H. Buller and G. Wilson (1999) *Implementation and Effectiveness of Agri-Environmental Schemes Established under Regulation 2078/92*. Project FAIR1 CT95-274, Final Consolidated Report, Volume 1, Institut fuer laendliche Strukturforchung (Research Institute for Rural Development), Frankfurt.
- Shaddish, W.R. and C.K. Haddock (1994) Combining Estimated of Effect Sizes. In: Cooper, H. and Hedges, L.V. (eds) *The Handbook of Research Synthesis*. Russel Sage Foundation, New York.
- Silvis, H.J. and C.W.J. van Rijswick (1999) *Tussen Interventie en Vrijhandel, WTO en de Nederlandse Agrosector*. (Between intervention and free trade, WTO and the Dutch agricultural sector). Report 5.99.05, Agricultural Economics Research Institute (LEI), The Hague.
- Soufi, R. and M. Tuddenham (1995) The Reform of the European Union Common Agricultural Policy. In: Gale, R., S. Barg and A. Gilles (eds) *Green Budget Reform*. Earthscan Publications Ltd, London.