TI 2000-083/3
Tinbergen Institute Discussion Paper

A Meta-Analysis of Environmental Impacts of Agri-Environmental Policies in the European Union

Katrin Oltmer
Peter Nijkamp
Raymond Florax
Floor Brouwer

A meta-analysis of environmental impacts of agri-environmental policies in the European Union

KATRIN OLTMER, PETER NIJKAMP, RAYMOND FLORAX
Free University Amsterdam, The Netherlands

FLOOR BROUWER
Agricultural Economic Research Institute, The Netherlands

Summary

This paper applies an ANOVA-type of meta-analysis to the evaluation of agri-environmental policy in the European Union. Meta-analysis is concerned with the statistical analysis of previous research results, and has become an established technique in the medical and natural science. The paper investigates whether specific conditions, under which agri-environmental measures are applied, have an effect on the behaviour of farmers with respect to three indicators, N-fertiliser, livestock density and area of grassland. The results indicate that agri-environmental policy intervention has a positive effect on the behaviour of farmers participating in agri-environmental programmes. The average premium per hectare and the absolute value of N-fertiliser, livestock density and area of grassland appear to have a significant influence on the behaviour of these farmers.

Keywords: agri-environmental policy, European Union, meta-analysis, environmental indicators, farmland preservation.

JEL classification: C19, D78, Q1, Q24

1. Introduction

Agri-environmental policy measures were introduced by the European Commission along with the MacSharry reform of the Common Agricultural Policy in 1992. This policy has been evoked, among others, by the increasing pressure of intensive agricultural production on the environment and by the changing role of land use in agricultural and rural areas during the 1980s. The function of farmland as the main input for agricultural production to secure food supply has gradually lost its importance. Instead, farmland as a provider of environmental amenities, such as wildlife habitats and scenic and cultural values for recreation purposes, has become increasingly public concern (Bromley and Hodge, 1990). These amenities can be regarded as positive externalities from traditional agricultural practices. However, increased intensification of modern production methods threatens the continuing existence of amenities. Amenities have the characteristics of a public good, and because of the lack of a market price, farmers' decision-making does not take their production into account (Brouwer and Slangen, 1998). Several studies document the public's willingness to pay for environmental amenities associated with agricultural and rural land (e.g., Drake, 1992; Pruckner, 1995; Kline and Wichelns, 1996; Brouwer and Slangen, 1998; Brunstad *et al.*, 1999).

The agri-environmental policy measure as formulated by the European Commission seeks to internalise the positive externalities from extensively used farmland, by offering compensation payments for the income losses farmers suffer when switching from intensive to extensive farming practices.[1] The compensation payments should obviously reflect the market price for the public good provided by farmers.

The participation in European agri-environmental programmes is voluntary for farmers. However, for the Member States it is obligatory to implement such programmes. It is thus the first common European framework for national policies in the agri-environmental field (Brouwer and Lowe, 1998). The agri-environmental policy measure is a very diverse and broad instrument that should be sufficiently flexible to consider 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 these policy takes place on a national, regional or even local level, resulting in a large number of different implementation strategies.[2] The problem for policy makers given this wide variety of implementation strategies is, that it is rather difficult to carry out cross-national comparisons of scheme effectiveness and that it is difficult to evaluate the economic efficiency of the schemes in general. Furthermore, in many cases the environmental policy target of the programmes is far too broad and not adequately identified, so that potential positive effects on the environment cannot be evaluated (Buller, 2000).

Against the above background, the present paper aims at offering a framework for comparative analysis of agricultural land use practices in various European countries. Agricultural land use practices are represented by three so-called environmental driving force indicators, which are 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 are related to specific policy measures, and others to general market or external conditions. This paper analyses drivers related to specific policy measures as well as those based on the structure of the agricultural sector.

Two major research questions regarding environmental aspects of agricultural land use are considered. One is concerned with the assessment of environmental effectiveness of agri-environmental policies in the European Union. Based on the perspective and the need to draw lessons from comparative case study research, the second research task of this paper deals with a methodological issue. The analysis will focus on whether meta-analysis is a suitable tool for policy assessment of agri-environmental initiatives in the EU.

The experience with agricultural policy in various European countries calls for a systematic research synthesis and comparison. We will employ 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; 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; Nijkamp and Baaijens, 2000). Meta-analysis aims to synthesise previous research findings or case study results with a view to identifying commonalities that might lend themselves to transferability to other, as of yet unexplored cases. The statistics of meta-analysis are in the mean time well-developed. Especially for quantitative case study results significant progress has been made. In this paper, we will use an ANOVA-type of meta-analysis adapted to effect size estimations in order to identify common drivers of agricultural dynamics in Europe.

The paper is organised as follows. Section 2 gives a short introduction regarding the use of environmental indicators in policy analysis and explains the environmental indicators used in our analysis. Section 3 presents the input data for the analysis, which originate from case studies of an EU-project. The methodology of meta-analysis and the statistical procedures applied in our analysis are introduced in Section 4. Section 5 reports on the results, and finally, Section 6 gives conclusions and recommendations.

2. 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. Firstly, they should provide information to policy makers and the general public about the state of the environment influenced by agriculture. Secondly, they have to help policy makers to better understand the cause-effect loops between agricultural activity and the environment. Thirdly, 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 societal groups 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 may become visible only after a number of years, or that the effects of agricultural production 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 3), since this project provides the input data for the meta-analysis. The FAIR research project developed 12 different indicators based. For the purpose of the meta-analysis a minimum amount of systematic and common data is needed. Since not all 12 indicators comply with this requirement, we were forced to deploy only three, namely mineral nitrogen fertiliser, livestock density, and grassland area per utilisable agricultural area. Our choice of indicators is hence solely based on practical reasons of 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, that can be summarised as follows.

- *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.

- *Livestock density:* A large amount of livestock per agricultural area is equivalent to high levels of manure and slurry. In turn, this is directly related to nitrate leaching into the groundwater. However, the actual relationship between livestock density has been found 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.

- *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. Finally, grassland is an ideal measure for the prevention of soil erosion through wind and water.[3]

3. Input data: Case studies of an EU project

The case studies used in the meta-analysis are the results of a three year project regarding the implementation and effectiveness of agri-environmental schemes in the European Union[4] (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 data collection took place through farm surveys based on a uniform questionnaire. Twenty-two case study areas were selected, two in each country, except for Sweden where four case study areas were selected. In total, 1000 farmers were interviewed, 50 in each case study area (and 25 in the Swedish case 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 criteria, 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 could also trace the environmental and socio-economic impacts of EU policies. For the purpose of this paper, we will mainly focus on the results of the environmental impact analysis (for a detailed description of this analysis we refer to Andersen *et al.*, 1999).

As mentioned in the previous section, three of the 12 environmental indicators developed by the research team have been selected for the analysis in the present paper. 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 interviewed farmers 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, 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 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 Student t-test and the Mann-Withney U-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. The structure of an experimental and a control group is a proper base for conducting a meta-analysis, where so-called effect sizes are calculated, which reflect the relative mean difference between these two groups. (See Section 4 for a detailed explanation of meta-analysis.)

Table 1 summarises the results of the change rate analyses carried out in the FAIR project for the three selected indicators. Expected and significant results are those where the respective change rates have the correct sign (negative for N-fertiliser and livestock density, and positive for grassland), and where the change rate of participating farmers is significantly larger than the change rate of non-participating farmers.

Table 1. *Results of the change rate analyses of the FAIR project.*

Indicator	… out of 22 CSAs	expected and significant	unexpected and significant	insignificant
N-fertiliser	9	- Sahagun (ES), 5% - Larisa (GR), 5% - Wetterau (GER), 10% - Cambrian Mountains (GB), 10%		- Devon Countryside (GB) - Rhoen (GER) - Viborg County (DK) - Vestjaelland (DK) - Nordburgenland (A)
Livestock	13	- Rhoen (GER), 5%		- Cambrian Mount. (GB) - Devon Countryside (GB) - Viborg County (DK) - Vestjaelland (DK) - Moura (P) - Castro Verde (P) - Nordburgenland (A) - Osttirol (A) - Schwarzwasser (CH) - Erlach/Seeland (CH) - Bocage-Avenois (F) - Wetterau (GER)
Grassland	13	- Enkoping (SW), 10%	- Devon Countryside (GB), 5%	- Cambrian Mount. (GB) - Viborg County (DK) - Vestsjaelland (DK) - Offerdal (SW) - Vallakra (SW) - Nordburgenland (A) - Osttirol (A) - Schwarzwasser (CH) - Erlach/Seeland (CH) - Rhoen (GER) - Bocage-Avesnois (F)

Table 1 shows that 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 13 results is significant for livestock density, and for grassland no significant result was found. However, it has to be kept in mind that the sample sizes in the individual case study areas tend to be rather small. This increases the probability of accepting the null-hypothesis that the average change rates of the two groups of farmers are not significantly different from each other although it may be false. This problem will be further elaborated in the following Section.

It should be noted that 9, 13 again 13 observations for the three indicators N-fertiliser, livestock density and grassland area are available for the analysis in this paper. Since we are bound by a limited number of observations, the analysis may be seen as a first exploration to apply the techniques of meta-analysis to agri-environmental policy evaluation. The statistical procedures of the meta-analysis employed in this paper are described in the following section.

4. Methodology of research synthesis

4.1. Introduction

Meta-analysis has a quite remarkable history in psychology and medical science and only recently found its way to economics. The development of meta-analysis in psychology and medical science is mainly due to the availability of a large amount of case studies on the same issue, performed in an experimental and largely standardised context, which forms a perfect base for statistical analysis. The lack of experimental and standardised conditions in many of the social sciences (including economics) is actually problematic when applying meta-analysis in a non-experimental context. In order to be able to compare existing research results in a strict statistical fashion, studies should be concerned with quantitative factors measured in identical units, or at least the results should be transformable into some common unit or index (Van den Bergh and Button, 1997).

Because of the quasi-experimental approach, the results from the case studies carried out in the FAIR project are suitable inputs for a meta-analysis. 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, which are in many cases insignificant. It was mentioned that this could be due to relatively small sample sizes. Standard statistical theory tells us that parameter estimates from large sample sizes are more robust than those from small sample sizes, because the variance of those estimates is smaller (Shadish and Haddock, 1994). Alternatively, estimates obtained from rather small samples are, due to their larger variance, 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 using simple vote-counting procedures (i.e., counting significant results only) is bound to lead to the conclusion that the average effect of the intervention is not significantly different from zero (Hedges and Olkin, 1985).

Meta-analysis artificially lowers the variance of the case study results, as it is based on the analysis of statistical summary indicators or effect sizes (e.g., the mean of the correlation coefficient) rather than on the original observations. The 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 4.2.)

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, notwithstanding that the analysis of the original data leads to results that are insignificant.

Meta-analytic techniques are also able to determine 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, there must be factors at work that are responsible for variations among the individual effect sizes. The identification of these factors is another task of the meta-analysis as it is carried out in this paper.

4.2. The effect size

Several differing definitions of effect size can be found in the literature. In the experimental sciences, two types of effect size are most commonly used: the *d*-type and the *r*-type. Examples of the *d*-type are Hedges' *g*, Cohen's *d* and Glass' *D*. An example of an *r*-type effect size is the correlation coefficient. Effect sizes of the *d*-type are generally standardised mean differences of control and experimental groups, which differ according to the way of standardisation.

The choice of which effect size measure to apply depends on the type of data, although in most cases the selection of a specific measure is not of crucial importance because many measures can be easily transformed into each other (Rosenthal, 1991). In our case, the original studies compare the change rates of two groups. They also report the means, standard errors and sample sizes of these change rates, and consequently it is most appropriate to calculate an effect size of the *d*-type. This analysis employs Hedges' g as the effect size measure. The interpretation of effect sizes of the d-type is as follows.

An effect size of the d-type reflects the difference in means between an experimental and a 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). The effect size can be interpreted as the z-score of the normal cumulative distribution function, where its respective $F(z)$-value is the proportion of control group scores that is less than the average score of the experimental group (Hedges and Olkin, 1985). 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% ($F(0.3) = 0.62$) of the individuals of the control group. Rosenberg *et al.* (1997) present Cohen's convenient rule of thumb about the interpretation of effect sizes: an effect size of 0.2 implies a small effect, of 0.5 a medium effect, and of 0.8 a large effect. Everything greater than 1.0 constitutes a very large effect.

4.3. Meta-analysis in four steps

The meta-analysis performed in this paper comprises 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. In the third step it is investigated whether the estimated effect sizes are homogeneous, which implies that 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, which is a moderator analysis.[5]

Step 1: calculation of the effect size

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

$$g = \frac{M_E - M_C}{S}, \tag{1}$$

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:

$$S = \sqrt{v} = \sqrt{\frac{(N_E - 1)V_E + (N_C - 1)V_C}{N_E + N_C - 2}}\,, \tag{2}$$

where V_E and V_C is the variance of the experimental and control group, and N_E and N_C the experimental and control group sample size, respectively.[6]

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. The most appropriate weight is the inverse of the variance of the respective effect sizes, as shown in the following equation (Shadish and Haddock,1994).

$$w_i = \frac{1}{v_i}\,, \tag{3}$$

where w_i is the weight and v_i the variance of the i-th effect size calculated according to Equation (2).

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

$$\overline{G_\bullet} = \frac{\sum_{i=1}^{k} w_i g_i}{\sum_{i=1}^{k} w_i}\,. \tag{4}$$

For testing the null hypothesis that the average effect size is not significantly different from zero, the Z-statistic will be applied. It is calculated as follows:

$$Z = \frac{\left|\overline{G_{\bullet}}\right|}{s_{\bullet}}, \tag{5}$$

where $s_{\cdot}$ is the average effect size standard error, that is given by:

$$s_{\bullet} = \sqrt{v_{\bullet}} = \sqrt{\frac{1}{\sum_{i=1}^{k} w_i}}, \tag{6}$$

and $v_{\cdot}$ is the average effect size variance. 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 effect.

Step 3: test on homogeneity of effect sizes

Equation (4) assumes that all individual studies share a common effect size. The test on the effect sizes of all individual studies not being homogeneous is called the *Q*-test, and it is given by the following equation:

$$Q = \sum_{i=1}^{k} \frac{\left(g_i - \overline{G_{\bullet}}\right)^2}{v_i} \tag{7}$$

If the value of Q exceeds the upper tail critical value of the χ^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 (4), therefore has 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 that critically influence 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 effect sizes among the individual studies. Another interpretation of moderator variables is that they 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 of a moderator of the first type is the premium level. Theoretically, it can be assumed that higher premium levels will induce larger changes in behaviour with respect to specific agricultural practice indicators. Secondly, there is the group of moderator variables that reflect the setting of the particular case study, such as country or time specific characteristics. Thirdly, there is a group of moderators that refer to methodological characteristics 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 present paper, the individual case studies all apply the same statistical technique. This means that methodological moderators are not relevant in our case.

The list of potential moderator variables is very long and the availability of information is the determining factor of which moderator variables to choose. The analysis in this current paper investigates the significance of the 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 should be related to larger effect sizes.[7]

II) Average farm size: With this moderator variable it is investigated whether effect sizes are influenced by the average farm size in the different case study areas.

III) Absolute level of indicator in 1997: Case study areas that in general have 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.

IV) Intensive versus extensive farming:[8] With this moderator variable it is investigated 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 variable it is tested whether effect sizes in areas of arable farming differ significantly from those in areas of husbandry farming.

The basic procedure in performing a moderator analysis is as follows. First, 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 can 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 equation:

$$Q_{between} = \sum_{i=1}^{p} \frac{\left(\overline{g_{i\bullet}} - \overline{G_{\bullet}}\right)^2}{v_{i\bullet}}, \tag{8}$$

where p is the number of groups, $\overline{g_{i\bullet}}$ the average effect size of the ith group, $\overline{G_{\bullet}}$ the overall average effect size (see Equation (4)), and $v_{i\cdot}$ the variance of $\overline{g_{i\bullet}}$, calculated according to Equation (6), taking into account the observations in that particular group only.

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

$$Q_{within} = \sum_{i=1}^{p} \sum_{j=1}^{m} \frac{\left(g_{ij} - \overline{g_{i\bullet}}\right)^2}{v_{ij}}, \tag{9}$$

where m is the number of observations in ith group, g_{ij} the jth effect size in the ith group, and v_{ij} its variance, according to Equation (6), taking into account the observations in that particular group only.

The sum of the Q-between and the Q-within statistic results in the overall Q-test applied to all observation (see Equation (7)):

$$Q = Q_{within} + Q_{between}. \tag{10}$$

In the ideal case, the selected moderator variable explains the heterogeneity in such a way that most of the heterogeneity occurs 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 group would still be large enough, a moderator analysis can be performed within the groups.

5. 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 5.1 describes the outcomes of Step 2 and 3. Section 5.2 gives the results of the moderator analyses.

5.1. Combined effect sizes and homogeneity test

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

Table 2. *Results of step 2 and step 3.*

	k	*N* (N_E, N_C)	*Hedges g*	*SE*	*Z*	*Q*	$P_{(Q)}$
N-fertiliser	9	349 (242,107)	-1.57	0.15	10.24*	52.24	0.00
Livestock	13	630 (445,185)	-0.82	0.11	7.35*	161.81	0.00
Grassland	13	569 (428,141)	-0.83	0.12	6.80*	169.84	0.00

The meaning of the symbols is as follows: k: number of case study areas; N: number of individual farmers; N_E: number of individual farmers in experimental group (participants); N_C: number of individual farmers in control group (non-participants); SE: standard error of Hedges' g.

Table 2 shows that the combined effect sizes of all three indicators are significantly different from zero. Although most of the original case studies show insignificant results, the combined effect sizes show 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, -1.57, which implies that 94% (F (1.57) = 0.94) of the change rates of non-participating farmers are lower than the average change rate of participating farmers. According to Cohen's rule of thumb (see Section 4.2), this reflects a very large effect of the policy intervention regarding the use of fertiliser. It should be noted that effect sizes can not be used to infer 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 as well.

The Q-test on homogeneity signifies at a very high significance level for all three indictors 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.

5.2. Results of the moderator analyses

Since the calculated effect sizes do not pass the *Q*-test on homogeneity, a moderator analysis as described in Step 4 is carried out. The moderators 'average premium per hectare', 'average farm size of participating farmers' and '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.

I) Average premium per hectare:

The results of the moderator analysis 'average premium per farm' are shown in Table 3.

Table 3. *Results of moderator analysis 'average premium per hectare'.*

	N-FERTILISER			LIVESTOCK DENSITY			GRASSLAND		
2 groups	*Hedges'g*	*Q*	$P_{(Q)}$	*Hedges'g*	*Q*	$P_{(Q)}$	*Hedges'g*	*Q*	$P_{(Q)}$
< 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	0.09		23.81	0.00		0.02	0.88
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	0.8	-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 groups comprise all case study areas where the average premium is less/larger than 40 ECU per hectare. For the indicators N-fertiliser and livestock density, the results are as expected, higher average premiums per hectare result into higher effect sizes. 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 as shown by the *Q*-between.

Since the *Q*-within tests in the 2-groups analysis still indicates heterogeneity among effect sizes, a second analysis was carried out, in which we tested whether a division in 3 groups improves the *Q*-within tests. The group division is indicated in Table 3. The table shows that only for the indicator livestock density increasing premiums per hectare result in higher effect sizes. The *Q*-between test still rejects the null hypothesis of homogeneity among the average effect sizes of the three different groups. The *Q*-within statistic slightly decreased, but there is still heterogeneity among the effect sizes within the groups. For the indicator N-fertiliser, the second group shows the largest effect size, and it is also one of the few cases where the *Q*-within test indicates homogeneity. For the indicator grassland, the *Q*-between test now signifies heterogeneity among the average effect sizes between groups. However, the unexpected negative signs 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 sizes. In addition to between-group heterogeneity, the *Q*-within tests should indicate homogeneity. This does not occur in this 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 Table 4.

Table 4. *Results of moderator analysis ‘average farm size’.*

	N-FERTILISER			LIVESTOCK DENSITY			GRASSLAND		
2 groups	*Hedges'g*	*Q*	$P_{(Q)}$	*Hedges'g*	*Q*	$P_{(Q)}$	*Hedges'g*	*Q*	$P_{(Q)}$
< 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	0.87		0.84	0.36		0.063	0.80
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	0.1	-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	0.29		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 with three groups. In the first analysis, the groups contain all case study areas where the average farm size of participating farmers is lower/higher than 80 ha. The *Q*-between tests of all three indicators signify homogeneity between the effect sizes of the two groups. This means that this 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.

The group division of the second analysis is presented in Table 4. 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 does not seem to be influential for the magnitude of the effect size. 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 heterogeneity among the effect sizes inside the groups in all cases. Unfortunately, the limited number of observations precludes a more differentiated analysis.

III) Average absolute value 1997

In the third moderator analysis, we divide the effect sizes of the different case study areas into two groups. For the indicator N-fertiliser, the groups contain those case study areas where the average absolute value in 1997 is lower/higher than 40 kg/ha. For the indicator livestock density, the groups comprise all case study areas with less/more than 1.5 livestock units per hectare on average in 1997. For the indicator grassland, the two groups are characterised by less/more than 50% grassland area per UAA in 1997. The results of the moderator analysis 'average absolute value in 1997' are shown in Table 5.

Table 5. *Results of moderator analysis 'absolute value 1997'.*

N-FERTLISER				LIVESTOCK DENSITY				GRASSLAND			
	Hedges g	*Q*	$P_{(Q)}$		*Hedges g*	*Q*	$P_{(Q)}$		*Hedges g*	*Q*	$P_{(Q)}$
<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	0.43
Q within		45.12	0.00			155.95	0.00			169.21	0.00

The *Q*-between test signifies heterogeneity between the average effect sizes of the two different groups for the indicators N-fertiliser and livestock density. This implies that the average absolute value in 1997 seems to have a significant influence on the magnitude of the average effect size. As expected, the case study areas with a higher absolute level of the indicator have a higher average effect. 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 heterogeneity among the effect sizes in all cases.

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-FERTILISER							
	Hedges g	*Q*	$P_{(Q)}$		*Hedges g*	*Q*	$P_{(Q)}$
Intensive	-1.49	11.90	0.01	Arable	-1.48	43.35	0.00
Extensive	-1.67	40.00	0.00	Husbandry	-1.87	7.68	0.01
Q between		0.34	0.56	Q between		1.21	0.27
Q within		51.90	0.00	Q within		51.03	0.00

Table 6 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 the case study area being characterised by an intensive or extensive, respectively, or an arable or husbandry agricultural production structure does not have an 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.

6. Conclusions and discussion

In this paper, the statistical methods of meta-analysis have been applied to agri-environmental policy evaluation in the European Union. Because of limited data availability, this study is rather exploratory. Nevertheless, some general conclusions can be drawn on the basis of this analysis. Firstly, the fact that meta-analysis artificially lowers the variance of the sample observations shows up in the results of Step 2, the combination of effect sizes. 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, implying that the change rate of non-participating farmers is lower than the average change rate of participating farmers. In general, the effect sizes of the indicator N-fertiliser show the highest value. This may be explained by the fact that the reduction of N-fertiliser is easier to organise and less dependent on other conditions as compared to 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 outweigh the payments for agri-environmental programmes. The effect sizes of the indicator share of grassland area per UAA unexpectedly exhibit negative signs. This paradoxical result 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 - like, 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 within the different groups. The occurrence of this problem does underline the diversity of the European landscape and the differences in the structure of the agricultural sector. This is often emphasised by researchers trying to evaluate European agri-environmental policy. The methodology of meta-analysis can shed more light on this diversity, if a sufficiently large number of observations (i.e., underlying case studies) is available. This would also make it possible to apply more advanced methods of meta-analysis that into account a set of moderator variables, for example multi-factor analysis or meta-regression analysis.

In the FAIR project it is suggested to introduce monitoring programmes with which the behaviour of participants and non-participants can be compared. In the context of such a quasi-experimental impact assessment, it would be better feasible to compare policy outcomes with policy objectives. Quasi-experimental case study results would also increase the number of potential input data for meta-analysis.

Notes

1. The agri-environmental policy of the CAP is formulated in the EC-Council Regulation 2078/92.
2. The agri-environmental policy includes about 2200 distinct measures incorporated in 127 programmes. 'Programmes' can be described as the way in which 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 (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 are discussed in CEC (1992), Buller (2000), and Deblitz (1999).
3. The mineral N-fertiliser indicator is measured in kg N-fertiliser per hectare. The livestock density indicator is measured in total livestock units (LU) per hectare of utilisable agricultural area. The grassland indicator is measured as a percentage of grassland per UAA.
4. Project FAIR 1 CT95-274 concerns the effectiveness of agri-environmental schemes established under Regulation 2078/92.
5. The description of the statistical procedure is based on Hedges and Olkin (1985), Rosenthal (1991, 1994) and Shadish and Haddock (1994).
6. 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, we use the adjusted, unbiased g, viz., g^u, that is obtained by applying $g^u=g\times c(m)$, where $c(m)=1-(3/(4(m)-1))$ and m is the degrees of freedom computed from the experimental and control group, (N_E+N_C-2). In the analysis c(m) is approximately 0.98, which means that the difference between g and g^u is rather small.
7. 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.
8. 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 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.

References

Andersen, E., Primdahl, J., Oñate, J., Aguirre, J., Cummings, C., Peco, B., Schramek, J., Klingler, B. and Knickel, K. (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, Institut fuer laendliche Strukturforschung, Frankfurt, 137-284.

Bergh, J.C.J.M. van den, and Button, K.J. (1997)
Meta-analysis of Environmental issues in regional, urban and transport economics, *Urban Studies* 34 (5-6): 927-944.

Biehl, D. (1999). Agri-environmental policies in a wider theoretical and institutional context. In J. Schramek, 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, Institut fuer laendliche Strukturforschung, Frankfurt, 45-49.

Bromely, D. and Hodge, I. (1990). Private property rights and presumptive policy entitlements: reconsidering the premises of rural policy. *European Review of Agricultural Economics* 17(2): 197-214.

Brouwer, F.M. and Lowe, P. (1998). CAP reform and the environment. In F.M. Brouwer and P. Lowe (eds.), *CAP and the Rural Environment in Transition.* Wageningen Pers, 13-38.

Brouwer, R. and Slangen, L.H.G. (1998). Contingent valuation of the public benefits of agricultural wildlife management: The case of Dutch peat meadow land. *European Review of Agricultural Economics* 25(1): 53-72.

Brunstad, J., Gaasland, I. and Vardål, E. (1999). Agricultural production and the optimal level of landscape preservation. *Land Economics* 75(4): 538-546.

Buller, H. (2000). The agri-environmental measure 2078/92. In: Brouwer, F. and P. Lowe (eds) *CAP Regime and the European Countryside: Prospects for Integration between Agriculture, Regional and Environmental Policies*. (forthcoming).

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.

Cooper, H. and Hedges, L.V. (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 Völkenrode, Sonderheft 195, Braunschweig.

Drake, L. (1992). The non-market value of the Swedish agricultural landscape. *European Journal of Agricultural Economics* 19(3): 351-364.

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

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

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

Kline, J. and Wichelns, D. (1996). Measuring public preferences for the environmental amenities provided by farmland. *European Review of Agricultural Economics* 23(4): 421-436.

Nijkamp, P., and Baaijens, S. (2000). Explanatory meta-analysis of tourist expenditures. *Journal of Policy Modelling*, (forthcoming).

Matarazzo, B., and Nijkamp, P. (1997). Methodological complexity in the use of meta-analysis for empirical environmental case studies. *International Journal of Social Economics* 34(719): 799-811.

OECD (1997). *Environmental Indicators for Agriculture*. OECD, Paris.

Petitti, D.B. (1994). *Meta-analysis, Decision Analysis and Cost-Effective Analysis for Quantitative Synthesis in Medicine*. Oxford University Press, New York.

Pruckner G.J. (1995). Agricultural landscape cultivation in Austria: An application of the CVM. *European Review of Agricultural Economics* 22(2): 173-190.

Rosenberg, M.S., Adams, D.C. and Gurevitch, J. (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: H. Cooper and L.V. Hedges (eds) *The Handbook of Research Synthesis*. Russel Sage Foundation, New York, 231-244.

Schramek, J., Biehl, D., Buller, H. and Wilson, G. (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 Strukturforschung, Frankfurt.

Shaddish, W.R. and Haddock, C.K. (1994). Combining estimates of effect sizes. In: H. Cooper and L.V. Hedges (eds) *The Handbook of Research Synthesis*. Russel Sage Foundation, New York, 261-281.