Understanding the Drivers of Farm Diversification in the Netherlands

A Master Thesis presented by

Manuela Meraner

15th July 2013

	University of Wageningen	University of Bonn
Registration number:	870519559130	2426122
Study programme:	Management, Economics and Consumer Studies	Food and Resource Economics
Chair group:	Agricultural Economics and Rural Policy Group / Regional Economics Group	Institute for Food and Resource Economics / Economic and Agricultural Policy Group
Course code:	AEP-81333	
Supervisors:	Prof. dr. W. Heijman Dr. T. Kuhlman	Dr. C. Wieck

Abstract

Farm diversification is a viable business strategy for farmers to react on increasing market pressure. In order to advise policy makers aiming to support multifunctional agriculture it is important to analyse the underlying drivers of farm diversification. This is done in this research by analysing diversification strategies using the Dutch Farm Structural Survey data from 2011. The study uses a binary logit model to determine the characteristics influencing the diversification decision in general. Furthermore, a more thorough analysis for six diversification strategies is conducted. Additionally the analysis categorises the specific diversification activities in order to estimate a multinomial probit model. This enables one to compare drivers leading to farm diversification in general, with drivers leading to specific activities. The analysis includes socio- demographic, economic and geophysical farm characteristics assumed to influence the diversification decision. The results show that the general diversification decision is supported by the same farm characteristics as most of the specific diversification activities. The largest differences in diversification drivers appear for characteristics influencing the probability of engaging in nature conservation. Consequently it is, important for policy makers to target policies at specific diversification activities in order to increase efficiency.

Table of contents

List of figures

List of appendix figures

List of tables

List of appendix tables

List of abbreviations

Arab	Arable farms
DU	Dunes
FC	Fluvial clays
LH	Löss hills
Hort	Horticulture farms
MC	Marine clays
Mix	Farms with mixed activities
Past	Pasture based livestock
PS	Pleistocene sands
EconSize	Economic size
Fem	Female workforce
FSS	Farm structural survey
FTE	Full time equivalent
IncDep	Income dependency
IIA	Independence of irrelevant alternatives
LEI	Agriculture Economics Research Institute
LR χ^2	Likelihood ratio chi-square
MNL	Multinomial logit
MNP	Multinomial probit
OECD	Organisation for Economic Co-operation and Development
PopDens	Population density
SO	Standardized average annual output

Introduction

Multifunctionality and diversification have been introduced within the European Common Agricultural Policy to escape the crisis of the productivist model of agriculture, where the emphasis was on raising farm output (Van der Ploeg and Roep, 2003, p. 37). In line with European policy makers one can argue that engaging in several on- and off- farm activities to increase the farm household's income is the future trend of European agriculture. Furthermore, diversified farms are seen to respond to upcoming consumer demands in a more flexible manner in order to maximize profits. A different interpretation of the farm diversification trend can be found in the simple need of farm holdings to survive in a tough market environment characterized by high variability in both yield and prices. In this line of argument, diversification is a return to the past in farming, as historically, farms used to pursue several activities simultaneously in order to ensure survival (Vries, 1993).

Heringa et al. (2012) present an extensive literature research on different definitions of multifunctional agriculture and its spill over effects on regional economic development in the Netherlands. However, there is still little known about the characteristics driving the farms' uptake of diversification activities. This thesis has the general objective to respond to the need for understanding farmers' characteristics leading to a diversification strategy. Determining the features of diversifying farms has significant implications for policy makers as certain socio-demographic farm characteristics as well as land or capital assets can directly be linked to the diversification decision. By helping to identify potential diversifiers the efficiency of policies aiming to support farm diversification can be increased. Additionally, the goal gains in relevance when considering numerous claims for a better understanding of the drivers of farm diversification and for an additional unfolding of theoretical groundings of rural development practices (Knickel and Renting, 2000; Van Der Ploeg et al., 2000; Wilson, 2007). For instance underlines Wilson (2007, p. 10): *"more empirical work will be needed in future to further substantiate theoretical and conceptual issues of multifunctional transitions"*. This thesis proceeds by firstly investigating the farm characteristics determining the general diversification decision, and secondly analysing the characteristics leading to specific forms of diversification. Thus, allowing a comparison of the drivers for specific diversification activities with the drivers for farm diversification in general.

The literature about the drivers of farm diversification strategies can be categorised into two streams. They are not mutually exclusive and can be interpreted as different viewpoints on one phenomenon; the societal viewpoint and the farm holdings viewpoint. From a societal viewpoint, one might interpret diversification as the internalisation of externalities (Finocchio and Esposti, 2008). This is supported by the argument that every farm holding produces, due to its multifunctional characteristics, positive and negative externalities. Furthermore, some of the positive externalities can be transferred into marketable goods, leading to an increase in farm income. From a farm holdings viewpoint the diversification of activities can be seen as a risk management strategy in an especially risky market environment (Aguglia et al., 2009; Mishra et al., 2004). Figure 1 illustrates the described levels of analysis, both leading to the diversification decision of the farm holdings.

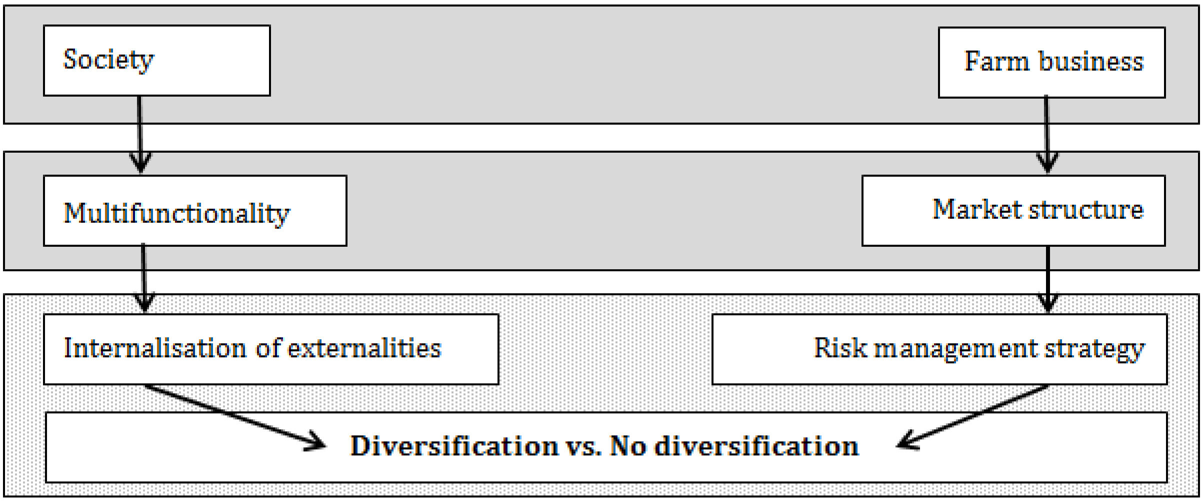

Figure 1 Different levels of diversification analysis
Source: own work

Diversification in this context is supported by rural development and aims to provide supplementary income. On the other hand, no diversification, i.e. specialisation on one activity, is supported by technological change and increases the farm's scale and/or margin. Moreover we define specialisation as non-diversification. This can go hand in hand with expansion of the farm business but can also be a 'business as usual' strategy, or imply focusing on one farming activity, i.e. concentration on one farm output.

The research is carried out using econometric methods to analyse Farm Structural Survey (FSS) data of the Netherlands of 2011. A binomial logit model is used to analyse the characteristics associated with the general diversification decision. Furthermore, the decision regarding the different diversification activities is analysed using a set of binomial logit models as well as a multinomial probit model.

The following section presents the theoretical background in which the essential distinction between multifunctionality and diversification is explained as well as some basic definitions and literature based theoretical implications of the analysed farm characteristics. Furthermore, follows the methodology used in the analysis. The next section describes the data set and descriptive statistics. The following short section gives some insights into the geographical clustering patterns of diversification activities. Subsequently the empirical model is explained. Following up is the presentation of the results. We then offer some conclusions, as well as recommendations for further research and policy makers.

Definitions and theoretical background

The concepts of diversification and multifunctionality of farm households are defined in various ways and interpreted in different contextual frameworks. The concept of multifunctionality in agriculture was introduced in 1992, where it was first mentioned within the context of the "United Nations Conference on Environment and Development" in Rio de Janerio. Multifunctional agriculture was here defined as: *"(...) multifunctional aspect of agriculture, particularly with regard to food security and sustainable development"* (UN, 1992). In 1998, the Organisation for Economic Co-operation and Development (OECD) gave the concept a more feasible shape by declaring that:

"Beyond its primary function of producing food and fibre, agricultural activity can also shape the landscape, provide environmental benefits (…) and contribute to the socio-economic viability of many rural areas. Agriculture is multifunctional when it has one or several functions in addition to its primary role of producing food and fibre." (OECD, 1998)

Within this framework, two main assumptions are made: 1) agriculture produces jointly commodity and non-commodity outputs, and 2) some of the non-commodity outputs have characteristics of externalities or public goods, hence leading to market failure. Furthermore, externalities are defined as positive or negative effects of production or consumption of goods and services on others. These effects are not accounted for in the production function on the supply side, or the utility function on the demand side, and hence are not reflected in the market price. This leads to a suboptimal situation. Positive externalities enhance social welfare and the absence of a market provokes an under-supply. Negative externalities on the other hand reduce social welfare and the absence of a market leads to an over-supply (Van Huylenbroeck et al., 2007).

The previous definition of multifunctionality and externalities leads to the conclusion that every farm household, independent of size, intensity and other characteristics, is producing intentionally or unintentionally, some commodity and non-commodity outputs. Van Huylenbroeck et al. (2007) propose that multifunctionality is a characteristic of the agricultural system in a certain rural area or region, and not necessarily of an individual farm. Whenever a farm holding makes the rational choice to internalize the positive or negative externalities, arising from the agricultural system's multifunctional characteristics, into the farm's production function we interpret this as farm diversification. This illustrates that contrary to the common use of farm diversification as a synonym for multifunctionality, both are different but strongly linked concepts (Wilson, 2007, p. 215). In the following we define diversification as: the reallocation and recombination of farm resources away from its original farming activity to generate another form of income (Ilbery, 1991).

Business management theory suggests that from a farm holding's point of view diversification can be interpreted as a risk management strategy. The increasingly competitive market environment in the agricultural sector has added to the risks the agricultural sector is exposed to: changing production conditions associated with changing weather conditions resulting in fluctuating yields; changing market conditions associated with changing prices or business cycles; and changes in policy (OECD, 2001). Empirical evidence shows an average annual decrease in the number of farm holdings in 2011 compared to the year 2000 of 2.9% in the Netherlands, this represents almost six farm closures per day. Furthermore the area of cultivated land in the Netherlands decreased in the same period on average by one per cent in total (Berkhout and Roza, 2012). Risk management is described by Kostov and Lingard (2003) as the economic behaviour of human beings when decreasing uncertainties through so called "risk defusing operators". Furthermore Kostov and Lingard (2003) interpret diversification as a risk avoidance strategy through a combination of "control" and "new activity" operators. Both of this risk defusing operators are not able to eliminate the risk completely but transform a risky environment into an environment with better assessed uncertainty and more knowledge. Diversification in this sense reduces the risk of volatile farm income by mitigating price risk and volatility in outputs since it reduces reliance on only one market and exposure to its price fluctuations (Robison and Barry, 1987).

Risk management is in this line of argument referring to the actor's attitude towards uncertainty not to the probabilities of its occurrence. It can be assumed that the farm's attitude towards risk is influenced by its characteristics the same way as the decision to which diversification activity the farm shifts its resources is. Thus treating all diversifying farms as one homogeneous group hinders one to distinguish differences amongst the drivers leading to different diversification strategies. Consequently a categorisation of diversification strategies is necessary. Ilbery (1991) suggests to distinguish two different types of diversification, "structural" and "agricultural" diversification. Structural diversification is focused towards the public and agricultural diversification is focused towards farming and different types of farm work. Barbieri and Mahoney (2009) present a more detailed concept, distinguishing several different types of diversification activities. Alternatively, Van der Ploeg and Roep (2003, p. 42) propose an operational classification of diversification activities into *deepening*, *broadening* and *regrounding*. The differentiation of diversification activities is pursued according to three sides of farming: first the agricultural side, second the rural side and third, the mobilisation of resources side (see figure 2). Deepening activities concern agricultural production, including those activities improving the product characteristics and or moving down the line of the supply chain (e.g. on- farm sale and processing). Broadening activities are related to the rural area side of the farm enterprise. Within this category, activities are characterised by creating new sources of income that are not related to agriculture (e.g. agro- tourism and care farms). The third side of farming is the mobilisation of resources; it is associated with activities involving off-farm labour or low-external input farming. These activities are referred to by Van der Ploeg and Roep (2003, p. 42) as regrounding of the farm process, and not considered in the following analysis.

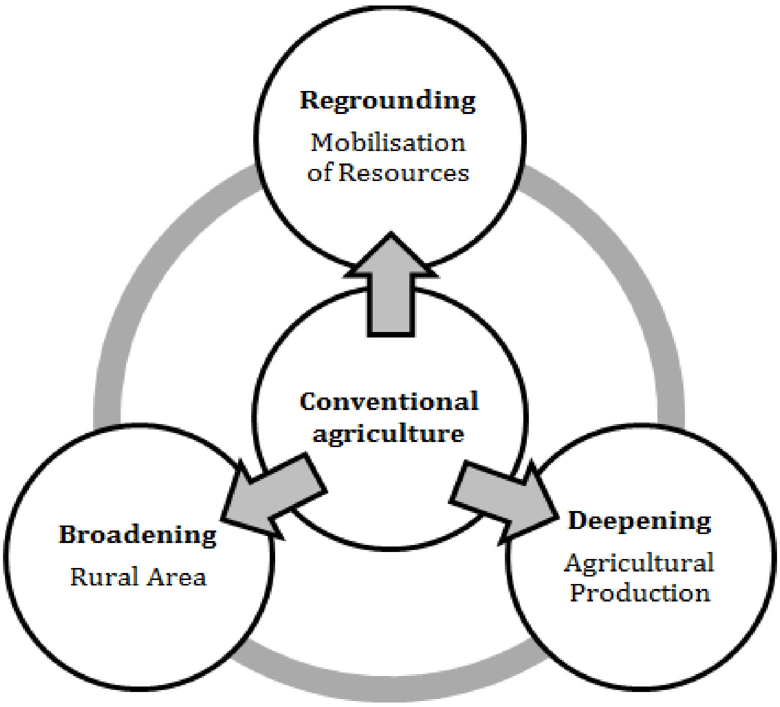

Figure 2 Classification of diversification activities
Source: own work adapted from: Van der Ploeg and Roep (2003, p. 43)

In 2007, the Dutch ministry for Agriculture, Nature and Food Quality (Ministerie van Landbouw, Natuurbeheer en Voedselkwaliteit) set up a taskforce for multifunctional agriculture (Taskforce Multifunctionele Landbouw) which aims to stimulate multifunctional agriculture in the Netherlands. The taskforce focuses on six diversification sectors: green care (health care in an agricultural setting), regional products (on-farm sale and processing), green services (nature management), tourism, child care, and education (Roest et al., 2009). The following analysis groups the diversification activities defined in table 1 according the categorisation introduced by Van der Ploeg and Roep (2003).

Table 1 Definition of diversification activities

	Definition
Deepening	
On- farm sale	Adding value to farm enterprises by direct marketing of agricultural products. This can be in the form of farm gate sales, farm shops, roadside stands, pick-your-own fruit and berry operations, or cut-your-own Christmas trees (Ilbery, 1991).
On- farm processing	Adding value to farm enterprise by processing cheese, butter, yogurt, cider/wine, jam/preserves, or craft activities (Ilbery, 1991).
Broadening	
Agro- tourism	Farm holdings that include a tourism component next to its farming activity. This includes the provision of accommodation and/or recreation in the form of camping, the hiring of a covered wagon, bicycle, water sports, catering, petting animals, or horse riding/stables (Dernoi, 1983).
Care farming	Farming practices aiming at promoting disadvantaged people's rehabilitation, education and care and/or towards the integration of people with 'low contractual capacity', but also practices that support services in rural areas for specific target groups such as children and the elderly (Di Iacovo et al., 2009).
Nature conservation	Support granted by the agri- environmental schemes in the Netherlands aiming at nature conservation, landscape maintenance and wildlife habitat creation. Including activities like meadow bird protection, delayed grass harvest and saving of landscape elements (Kleijn et al., 2001).

After having identified and classified the main diversification types this study will focus on, it is necessary to take a closer look at the drivers leading to the different diversification activities. Ilbery (1991) suggests that diversification is the outcome of "internal" and "external" factors. We capture some of those factors based on a thorough literature research with three categories of farm characteristics: socio-demographic characteristics, economic characteristics and geographical characteristics.

A number of studies analysing farm characteristics have shown that socio-demographic characteristics of the farm household, including the main farm operator's age, and the availability of female workforce influences the diversification decision significantly. Pope and Prescott (1980) suggests that younger farmers are more risk averse because they are less wealthy, this corresponds with Barbieri and Mahoney (2009) who conclude that larger long term ties for younger farmers increase the need to strengthen the farm business through diversification.

The availability of female workforce gives indications on the farm family's size and labour endowment of the farm holding. Mishra et al. (2004) suggest a stronger need to create employment on the farm for larger families leading to more diversification amongst larger families. Nilsson (2002) finds, in a qualitative study of farm diversification drivers, that the rather traditional division of work in most farm households between male and female tasks leads to an increasing uptake probability of diversification activities with increasing availability of female workforce. He reveals that the farm women look for additional occupation while the farm men are out in the fields. Moreover Hjalager (1996) proposes that especially broadening activities benefit from available female workforce as they imply making tasks performed by female household members profitable for the farm holding.

The farm size has been shown to be of significant influence on the diversification decision by numerous researchers. Some studies suggest that larger farms are more likely to be diversified, since they can allocate and exploit available resources more efficiently (Ilbery, 1991; Pope and Prescott, 1980). However, based on Krugman (1990), Mishra et al. (2004), and Vik and McElwee (2011) concluded that larger farms are more specialized based on the perceived benefits from economies of scale.

The income share retrieved from the diversification activity gives insights into the dependency of a farm holding on the diversification activity. Activities that require larger resource shifts away from the main farming activity can be assumed to be associated with higher returns and consequently higher income dependencies. Hjalager (1996) shows in an interview based study amongst tourism farms in Denmark that larger scale farms are able to invest more into diversification activities, and consequently expect a higher additional income.

Several papers find that diversification is also associated with the farm type. Amongst others, McNally (2001) retrieves that the seasonality of the farming activity determines the diversification decision. More seasonal farming activities, like arable/crop farming, are assumed to have a positive influence on the diversification decision since off season they have spare time to develop diversification activities. Contrarily, farming activities with constantly high labor intensity are assumed to have less excessive time to devote to a diversification activity. Jongeneel et al. (2008) on the other hand argue that that on-farm sale is not attractive for arable farms due to its time intensive nature and low remuneration per hour.

The literature distinguishes two ways in which geographical characteristics influence the diversification decision: firstly the attractiveness of the landscape in which it is located, and secondly the impact of urban areas. Mishra et al. (2004) and Pfeifer et al. (2009) find that the soil productivity influences the cropping and diversification decision. They conclude that farms on less productive soils are exposed to a higher farming risk and more volatile yields, consequently they are assumed to increasingly look for additional income outside the main farming activity. Ilbery (1991) shows that the proximity to main roads and urban hubs increases the probability to uptake a diversification activity. He reasons that the better access to the demand side makes marketing and customer binding more efficient. Barbieri and Mahoney (2009) and Mishra et al. (2004) show on the other hand that farm diversification is negatively influenced by closeness to urban areas. This is consistent with land pricing literature explaining that farms located closer to urban areas are more likely to choose an exit strategy or to look for off- farm employment due to higher returns on capital and labour outside the farming business (Lange et al., 2013).

Inspired by researchers like Pfeifer et al. (2009) and based on “new economic geography” theory we additionally analyse clustering patterns of diversification activities in the Netherlands. Baldwin and Wyplosz (2004) explain that geographical concentration of activities is a result of agglomeration and diffusion forces. Pfeifer et al. (2009) find that the proximity to national parks has a positive influence on the clustering of combinations of diversification activates, due to positive spill over effects of complementary activities.

Furthermore one can assume that for nature conservation activities the content of Dutch support measures, influences the uptake probability. Kleijn et al. (2001) find that the most important Dutch agri-environment schemes are meadow bird agreements. Thus a correlation of the preferred meadow bird habitat and increased nature conservation participation is likely. For a more detailed analysis of the influence of the farmers' attitude towards the government, on the participation in agri-environment schemes consider Jongeneel et al. (2008).

The selection of driving characteristics is determined by the availability of information in the FSS database and previously presented literature. The choice of farm characteristics analysed in this research, as well as their measurement are depicted in table 2. Each variable is assigned to a specific -more detailed- research question aiming to make the overall goal of understanding the driving characteristics of farm diversification activities more feasible.

Table 2 Explanatory variables with associated research questions

		Research question	Literature supporting the research question
Socio demographic characteristics			
Age Age of the main farm operator.	**RQ1**	Are younger farmers more likely to engage in broadening activities?	Finocchio and Esposti (2008) Jongeneel et al. (2008)
Availability of female workforce Amount of female workers in full time equivalents (FTE) available.	**RQ2**	Are farm households with more female workforce more likely to engage in agro-tourism and care farming activities?	Nilsson (2002) Hassink et al. (2007) Hjalager (1996)
Economic characteristics			
Farm size Economic farm size measured in Standard Output units (SO).	**RQ3**	Are large farms more likely to specialise?	Mishra et al. (2004)
Income dependency indicator Percentage of farm income retrieved from diversification activity.	**RQ4**	Are the farms that mainly rely on income from diversification activities mainly agro-tourism farms?	Hjalager (1996)
Farm type Farm types according to the main income source.	**RQ5**	Are farms with arable specialisation less likely to engage in on farm sale activities?	Jongeneel et al. (2008)
Geographical characteristics			
Urbanisation Population density, measured as inhabitants per km^2 on municipality level.	**RQ6**	Are direct sale activities more likely to be situated in urban areas?	Ilbery (1991)
Soil Categorisation of six soil types predominant in the Netherlands.	**RQ7**	Are nature conservation activities more attractive on less productive soils?	Pfeifer et al. (2009)

Methodology

Within the theoretical framework of the utility maximisation model we assume that the diversification decision is based on the rational choice of each farm holding. Additionally we assume that the decision maker has perfect discrimination capability between several risk management choices. This implies that the optimal risk management strategy chosen by each farm reflects its utility maximizing option. And inevitably leads to the conclusion that the observable diversification choices are always the optimal ones. As the true utility function cannot be observed directly we assume that the observable optimal choice is a linear function of socio- demographic, economic and geographic farm characteristics.

General diversification model

The general diversification decision can be interpreted as a binary choice model. The latent utility difference between diversification and no diversification y_i^*, is assumed to be determined by a linear function of observed characteristics plus an unobservable error term (ε_i) (Verbeek, 2008, p. 203):

$$y_i^* = \beta_i x_i + \varepsilon_i \qquad \varepsilon_i \sim Logistic\ (0,1)$$

$$\text{with } y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* < 0. \end{cases} \tag{1}$$

Where x_i represents a vector containing socio–demographic, economic and geographical characteristics. The error ε_i is assumed to follow a standard logistic distribution. The probability that the observable dependent variable y_i is one equals the probability that the utility difference is positive (for a more detailed explanation of the model and comparison with the alternative binomial probit model see appendix C).

Specific diversification activity model

The probability to adopt a specific diversification activity is first modelled as a separate binary choice model for each activity. They follow the same theoretical framework as presented in the previous section (see equation (1)). y_i^* describes here the unobserved difference between the utility gained from the specific activity and the utility gained from any other diversification activity.

Furthermore, a multinomial choice model is used to estimate simultaneously utility differences among the previously introduced diversification categories: broadening ($j = 1$), deepening ($j = 2$), combined activities ($j = 3$), and no diversification ($j = 0$). The model is constructed using a latent variable indicating the difference in utility gained from each possible category (Verbeek, 2008, p. 223):

$$y_{ij}^* = \beta_{ij} x_{ij} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\ (0, \Sigma) \text{ and } j = (1 \ldots 3)$$

$$\text{with } y_i = \begin{cases} 1 \text{ if } y_{i1}^* > 0 \\ 2 \text{ if } y_{i2}^* > 0 \\ 3 \text{ if } y_{i3}^* > 0 \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

Where β_{ij} is a vector of parameters specific to the j-th alternative associated with the vector x_{ij} which contains the observable farm's characteristics. The error terms ε_{ij} are assumed to be multivariate normally distributed with mean zero (for a specification of the covariate error term and comparison with the multinomial logit model see appendix C).

Data and descriptive statistics

The analysis is based on data from the FSS, collected by the Dutch authorities and processed by Statistics Netherlands. The data is provided by the Dutch Agricultural Economics Research Institute (LEI) and consists of information collected from 70,392 farms in the year 2011. The data cleansing process leaves a sample consisting of 68,724 farms. The excluded data consist of farms which stated to engage in none of the diversification activities proposed, but entered a positive percentage of income retrieved from the diversification activity. One reason for this mismatch can be assumed to be a general misunderstanding of the question or differing definitions of "diversification activity". The missing data is smaller than 10%, and can consequently be ignored (Hair et al., 2010, p. 49). The FSS covers all farms in the Netherlands, and is thereby representative over time for Dutch agriculture. Thus, the general reliability of data and the data gathering process can be considered as good.

We will now discuss each of the farm characteristics and their likely influence on the diversification decision in more detail. The results from the descriptive statistic are depicted in appendix B in table B.1 – table B.4.

Socio demographic characteristics

Age

The age of the main farm operator has been shown to play an important role in the farmer's attitude towards risk, it is included in the analysis as '*AGE*'. The average age of farm operators is 54 years. The average diversifying farmer is about two years younger than the average non-diversifying farmer. On- farm processing and care farmers have the youngest farm operators, whereas tourism farmers have the oldest farm operators amongst diversifying farmers.

Female workforce

The average farm household has about one female worker on the farm measured in full time equivalents (FTE). In the following the average female workforce available is indicated as 'Fem'. Not-diversifying farms have on average fever female workforce available. Care farms and processing farms have the largest numbers of female workers, farms engaging in nature conservation have the fewest.

Economic characteristics

Economic size

The explaining variable economic size is measured in standardized average annual output (SO). In the following regression it is referred to as '*ECONSIZE*'. The analysis divides the SO figures by 100,000 to yield interpretable results. Actual output per farm is not known, but SO is defined as the average annual agricultural output, calculated over all farms, measured in Euros, per hectare or animal generated by the crop or animal category on the specific farm (Berkhout and Roza, 2012). Farm payments and subsidies are not included in the SO. For the year 2011 the price level of 2007 is applied as base. The average farm yields about €280,000 annual output.

Diversifying farms yield on average €85,000 less annual agricultural output than not diversifying farms. Agro- tourism farms yield the lowest average annual output, whereas processing farms are on average the largest of all diversifying farms.

Figure 3 shows the distribution of each activity within a certain SO category in €1,000. Diversification activities are decreasing amongst farms with an SO equal to or larger than €400,000. With increasing economic size less farms uptake a diversification strategy, this is corresponding with the theory that diversification is used to reduce the risk associated with farming, by spreading their activities.

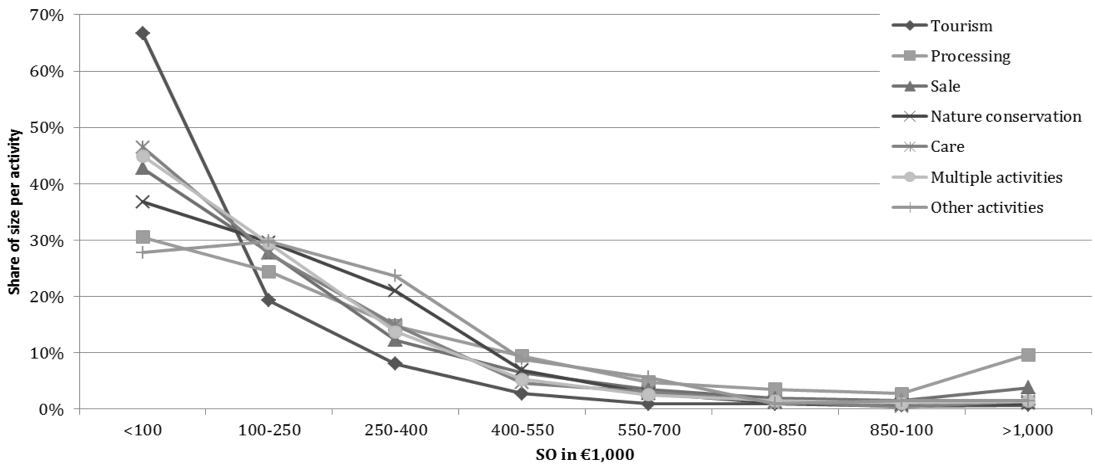

Figure 3 Size distribution of diversification activities

Amongst on- farm processing and sale farms one can see a small upward trend in diversification activities for farms larger than €1,000,000. Below €100,000 SO one finds 31% of all processing farms, this is steadily decreasing with increasing farm size, but then starts to increase again to a 10% share of processing farms larger than one million Euro SO. The same pattern, but less strong, is found for on- farm sale: 43% of sale farms are smaller than €100.000. This share is decreasing until a SO larger than one million, where we find an increase (4% of all sale farms). Within the share of tourism farms, the majority (67%) are smaller than €100,000 SO, which is so small that the income dependency on agro- tourism must be high for the farm business to survive. One has to note that the reported size indicates the average agricultural output not including income retrieved from diversification activities.

Farm type

This study uses the European Commission's typology of farms with its categorisation of different farm types according to the relative income distribution coming from different production sources (European Commission, No 1242/2008). The data on farm types covers several dozens of different types. For the ease of understanding the study groups some farm types together. The selected types are included as dummy variables referred to as following: mixed farming '*MIX*' (mixed cropping, mixed livestock, and mixed crops/livestock holdings), pasture based livestock '*PAST*' (grazing livestock), intensive livestock '*INT*', horticulture '*HORT*', and arable farming '*ARAB*'.

Intensive livestock and horticulture diversify less than average; mixed farms diversify the most. The other subsectors are close to average. Within the diversifying farms, 63% are pasture based livestock farms, 12% are arable farms, 13% are horticulture farms. Mixed farming activities presume a share of 7% of all diversifying farms and 3% of all diversifying farms' main farming activity is based on intensive livestock breeding.

Income dependency

The share of dependency on the income retrieved from diversification is in the regression labelled '*INCDEP*'. More than half of all diversifying farm holdings state to retrieve less than 10% of their income from the diversification activities (see figure A.2). Diversification activities with relatively small capital investments, like nature conservation, are associated with relative lower income dependencies.

<u>*Geographical characteristics*</u>

The impact of urban areas is incorporated by using the population density on municipality level as a proxy for the urbanisation degree. The attractiveness of the landscape in which it is located is incorporated in the analysis with six different soil types, predominant in the Netherlands. Additionally we connect the spoil type with clustering patterns of the diversification activities.

Urbanisation

In order to measure different degrees of urbanization the population density is chosen as a proxy. The population densities reported on municipality level are connected to the farm holdings in the data set. Thus every farm can be associated with a population density measured in inhabitants per km^2. The variable is divided by 100 to make estimated results more feasible, it is in the following referred to as '*POPDENS*'. Care farming, on- farm sale and processing activities are on average situated in more urbanised areas. The opposite holds for nature conservation and agro- tourism farms.

Soil type

In the Netherlands one can distinguish six different main soil types, which are incorporated in the study as dummy variables: dunes '*DU*', löss hills '*LH*', Pleistocene sands '*PS*', peat '*PE*', fluvial clays '*FC*', and marine clays '*MC*'. Each farm in the sample is assigned to the soil type it is mainly situated at. Most farms in the Netherlands are situated on Pleistocene sands, the predominant soil type across the country.

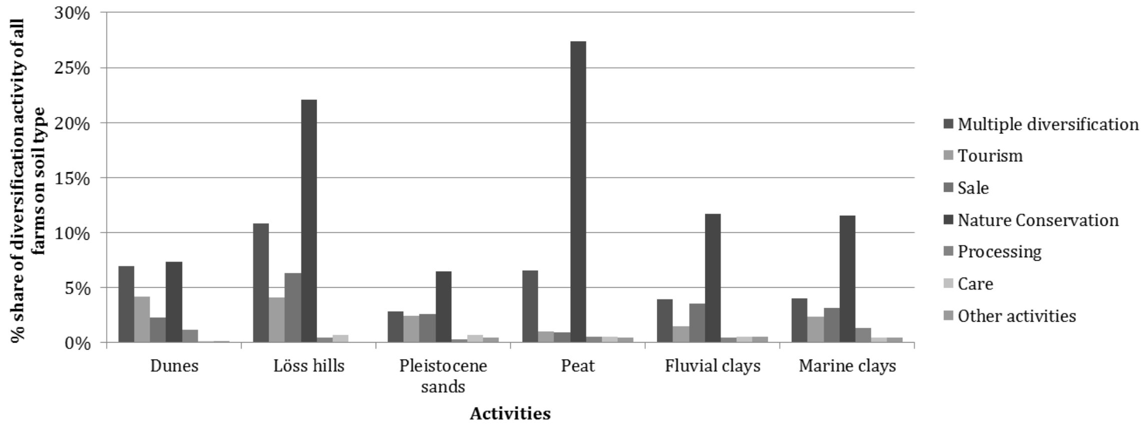

Figure 4 Distribution of diversification activities on soil types

In figure 4 one can see the distribution of diversification activities amongst the different soil types. It shows the share of each diversification activity on the total number of farms on a specific soil type. Overall one can find the most diversification on löss hills (44%), and peat (37%). Next to the largest share of nature conservation activities amongst all soil types, one can see that at the coastal areas, on dunes, multiple activities are largely represented. On löss hills, which are mainly to be found in the south of the country, multiple activities and sale have the largest share next to nature conservation. On Pleistocene sands, the diversification activities are distributed relatively equal. Peat soils show the largest share of nature conservation farms (27%).

Clustering patterns of diversification activities

A common assumption in "new economic geography" is the influence of dispersion and agglomeration forces. Agglomeration forces are assumed to lead to geographical concentration, or clustering, of similar businesses in a region whereas dispersion forces tend to diffuse business structures in a region (Baldwin and Wyplosz, 2004, p. 280). Figure 5 b) depicts the darker shaded diversification density across the Netherlands. As one can see there are big clusters of diversification activities found in the north and in the north- west of the country. In figure 5 a) one can see the distribution of the six different soil types in the Netherlands. When comparing the maps in figure 5 one can see, that the largest diversification clusters are found on peat soils, this corresponds largely with the findings for the largest group of diversification activities, nature conservation (see figure 6). Additionally we find clustering of diversification activities around the very light area in the centre of the country, where a big national park is located.

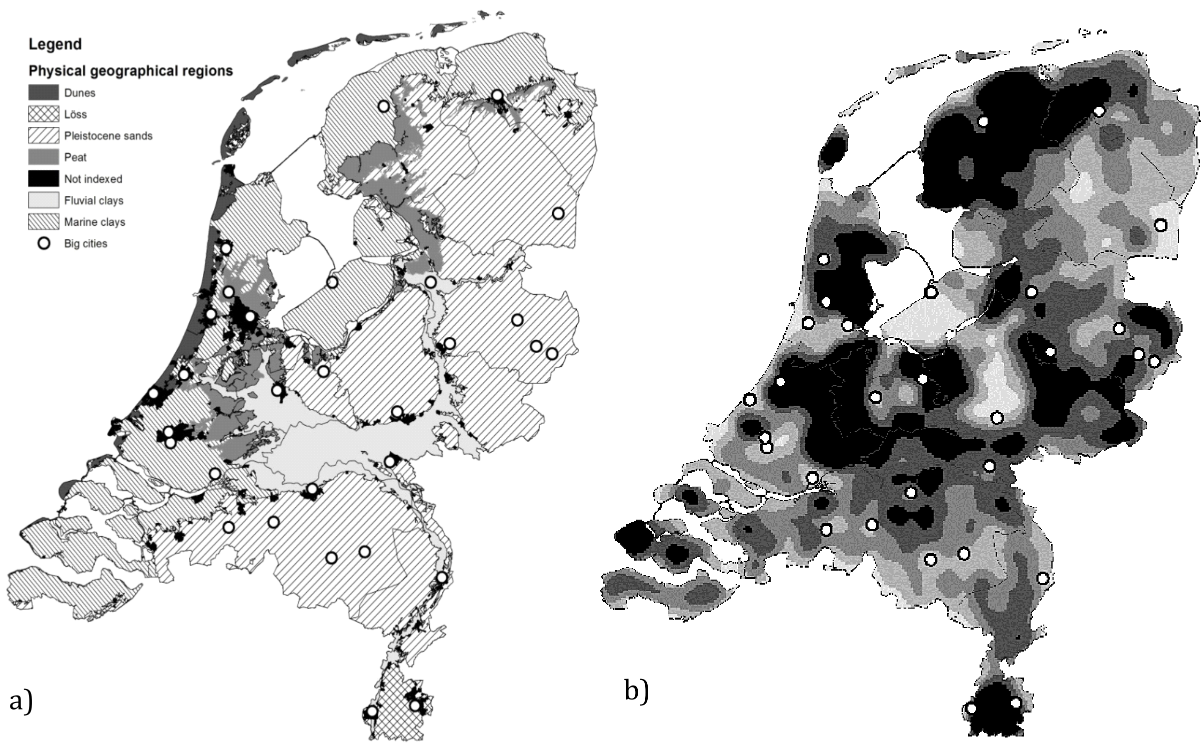

Figure 5 a) Soil type distribution in the Netherlands, b) Density of diversification activities
Source: own work based on data from the Dutch FSS 2011

In figure 6 one can see the density of different diversification activities in the Netherlands. On- farm sale activities are especially clustered in the southern area of the country; this is corresponding with the findings for the soil type, and a correlation with löss hill areas. Agro-tourism activities show to be clustered in the coastal areas but also in the west of the country.

Care farms do not follow a clear clustering pattern, indicating that this activity is influenced more by dispersion forces. Processing activities show clear geographical concentration in the west- and north-west areas of the Netherlands. Multiple diversification activities show clear clusters for the same areas as nature conservation and tourism, which is the most common combination of multiple diversifiers.

Figure 6 Density of diversification activities in the Netherlands
Source: own work based on data from the Dutch FSS 2011

Empirical model

To analyse the complex motives for farm diversification this research is laid out in two parts: The first part of the study analyses the characteristics determining the decision whether or not to diversify. A binary logit regression is used to analyse the probability of diversification uptake. The second part of the study focuses on the analysis of characteristics determining the engagement in a specific diversification strategy. For this purpose the five main diversification activities introduced in table 1 are selected and a binary logit regression is carried out for each of them. Consequently, only the farms that choose to diversify their activity in any of the selected ways are considered in this analysis. Furthermore, the specific activities are grouped according to their diversification category (see figure 2) and analysed using a multinomial probit model (MNP). For the analysis the statistical software package STATA/IC 10.1 is used.

General diversification decision

The general diversification decision is estimated using a binomial logit regression model. The latent variable $Diversification_i^*$ describes the difference in utilities gained from diversification and no diversification. It is assumed to depend on the selection of 13 observable farm characteristics and an unobservable error term ε_i. The estimated specification of the model has the following form:

$$\begin{aligned} Diversification_i^* \\ &= \beta_0 + \beta_1 Age + \beta_2 Fem + \beta_3 EconSize + \beta_4 Past + \beta_5 Arab + \beta_6 Hort \\ &+ \beta_7 Mix + \beta_8 PopDens + \beta_9 Mc + \beta_{10} Ps + \beta_{11} Fc + \beta_{12} Du + \beta_{13} Lh + \varepsilon_i \end{aligned} \quad (3)$$

$$\text{with } Diversification_i = \begin{cases} 1 & \text{if } Diversification_i^* > 0 \\ 0 & \text{if } Diversification_i^* = 0. \end{cases}$$

The McFadden R^2 is rather low (0.0546) which gives rise to the assumption that the uptake of a diversification strategy is hard to explain by the existing variables. The χ^2 of the model is significant indicating that all variables are jointly significantly different from zero. The cross-tabulation of actual and predicted outcomes reveals that about 80% of the estimated probability outcomes are correctly predicted in the model. The likelihood ratio χ^2 test confirms significance of all variables at the 10% level, apart from the dummy variable for the farm type horticulture. Testing for multicollinearity, resulted in no approximate relationship amongst the parameters (for a more detailed explanation of the overall goodness-of-fit measures and multicollinearity test see appendix C).

Specific diversification activity decision

To analyse the characteristics increasing the probability of engaging in a specific diversification activity. We introduce an additional variable representing the dependency of a farm's income on the diversification activity.

Firstly, a set of binary logit models is estimated. Here each specific activity is compared to the general diversification sample including 13,894 farms. In order to simplify the analysis, the system of binomial logit models depicts only the diversification activities of most interest for Dutch policy makers as defined in table 1, and groups the rest (i.e. aquaculture, agricultural kindergartens, and education on farms) as 'other activities'. Additionally the analysis requires to include the farms choice to engage in more than one activity simultaneously as a separate group (see appendix A). The different activities are distributed as following: multiple diversification (18%), tourism (11%), sale (14%), nature conservation (49%), processing (3%), care (3%) and other diversification activities (2%).

Secondly the diversification activities are grouped according to the three categories introduced earlier: broadening (15%), deepening (4%), and combined activities (2%). These categories are used to estimate a multinomial probit model where no diversification (79%) forms the baseline outcome. Equation (3) is specifically estimated as:

$$Diversification\ category^*_{ij} = \begin{cases} y^*_{broadeing} & = \beta_{i1}x_{i1} + \varepsilon_{i1} \\ y^*_{deepening} & = \beta_{i2}x_{i2} + \varepsilon_{i2} \\ y^*_{combination\ of\ activities} & = \beta_{i3}x_{i3} + \varepsilon_{i3} \end{cases}$$

$$\text{with } Diversification\ category_i = \begin{cases} 1 \text{ if } y^*_{broadeing} & > 0 \\ 2 \text{ if } y^*_{deepening} & > 0 \\ 3 \text{ if } y^*_{combination\ of\ activities} & > 0. \\ 0 \text{ if no diversification.} \end{cases} \tag{4}$$

Consequently this data set includes all farms again. The Wald χ^2 (42) test confirms the significance of the multinomial probit model (for a more detailed explanation of the significance tests applicable to MNP see appendix C).

Results

The interpretation of a logistic model is somewhat different and more complex compared to a linear regression model. Table 3, table 4, and table 5 report the estimated marginal effects on the probabilities of each variable with the form $\frac{dy}{dx}$. For continuous independent variables, the marginal effect measures the change of probability given a one unit change of the independent variable, holding all other variables at their means; this may result in very small values (Verbeek, 2008, p. 202). For dummy variables a change from 0 to 1 leaving all other variables constant at their mean is reported.

General diversification model

Table 3 Estimated marginal effects of the general diversification model

General diversification model	
Socio- demographic characteristics	
Age	-0.31***
Fem	1.15***
Economic characteristics	
EconSize	-0.64***
Farm type	
Past	11.92***
Arab	4.64***
Hort	-0.77
Mix	17.62***
Geophysical characteristics	
PopDens	-0.08**
Soil type	
MC	-7.46***
PS	-17.77***
FC	-8.81***
DU	-4.54***
LH	7.49***

*** Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level**

Socio- demographic characteristics

The general model estimates age to have a significant negative effect on the probability to diversify. The probability to diversify decreases with each additional year of the main farm operator by 0.31%. The number of female workforce available on the farm is estimated to have a significantly positive impact on the likelihood to diversify. With one more woman working on the farm the diversification probability increases by 1.15%.

Economic characteristics

The economic size measured in SO, has a negative effect on the diversification probability. The probability to diversify decreases by around 1% with an increasing SO by one unit. Consequently research question three: 'Are large farms more likely to specialise?' can be answered positively, keeping in mind that we defined specialisation as non-diversification. With increasing farm size the propensity to diversify decreases, and the propensity to specialise increases, confirming the results from Mishra et al. (2004).

The farm type as an indicator influencing the probability to diversify shows significant results for all included types except the estimates for horticulture farms. Intensive livestock farms are used as reference category. Farms with a mix of farming activities are the most likely to diversify. The probability to diversify is amongst mixed farms around 18% higher than the probability to diversify for intensive livestock farms. Pasture based farms also show a high probability to choose a diversification strategy.

Geophysical characteristics

With increasing population density, by one person per km^2, the probability to diversify is decreasing by 0.08%. The soil type where the farm is situated at shows to significantly influence the diversification probability for the six soils predominant in the Netherlands. Peat is here the reference group for the dummy variables. Amongst farms on löss hills probability to diversify increases by 7.49%. All other soil types show a negative probability to diversify, in other words farms on the other soil types are less likely to diversify then farms on peat soils.

Specific diversification activity decision

Table 4 Estimated marginal effects of six specific diversification activities

Diversification activity specific models						
	Tourism	Sale	Nature conser-vation	Processing	Care	Multiple diversi-fication
Socio- demographic characteristics						
AGE	0.03*	-0.03	0.41***	-0.02***	-0.05***	-0.20***
FEM	0.83***	-0.03	-12.57***	0.03	0.18***	1.31***
Economic characteristics						
ECONSIZE	-1.59***	-0.17***	-0.92***	0.12***	-0.11**	0.05
INCDEP	2.85***	-1.13***	-17.35***	0.45***	0.75***	6.58***
Farm type						
PAST	-3.21***	-19.59***	33.31***	-1.60**	-2.09***	-3.16*
ARAB	-3.00***	-1.56**	18.64***	1.98**	-2.38***	-6.58***
HORT	-6.64***	15.89***	-32.35***	2.17**	-1.81***	-2.12
MIX	-4.53***	1.16	3.48	0.49	-1.20***	3.39
Geophysical characteristics						
POPDENS	0.02	0.07*	-0.43***	-0.01	-0.02	0.01
Soil type						
MC	15.53***	7.86***	-17.66***	0.70	0.98	-0.81
PS	17.23***	10.32***	-25.80***	-0.96***	2.33***	-1.92*
FC	9.95***	11.74***	-12.30***	-0.55*	1.45*	-1.04
DU	32.07***	2.21	-26.58***	0.72	-1.32**	9.92***
LH	14.36**	5.76**	-7.92***	-1.14***	0.20	4.46**

*** Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level**

Table 5 Estimated marginal effects of three diversification activity categories

Diversification category model			
	Broadening	Deepening	Broadening and Deepening
Socio- demographic characteristics			
AGE	-0.04***	-0.01***	0.00***
FEM	0.15**	0.04***	0.06***
Economic characteristics			
ECONSIZE	0.03	-0.03***	-0.05***
INCDEP	14.57***	1.46***	3.64***
Farm type			
PAST	8.13***	0.01	-2.62***
ARAB	-2.94***	-0.62***	-1.11***
HORT	-8.03***	0.30*	2.45***
MIX	2.39***	1.09***	1.26***
Geophysical characteristics			
POPDENS	-0.07***	0.00	0.00
Soil type			
MC	-4.78***	-0.30***	1.13***
PS	-10.19***	-0.53***	0.59**
FC	-5.10***	-0.26***	0.87**
DU	-1.78**	0.17	-0.14
LH	1.06	0.79**	2.13***

*** Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level**

Socio- demographic characteristics

A one year increase in age has a positive effect on the probability to engage in tourism and nature conservation. The probability to engage in on- farm sale, processing, care farming and multiple activities is negatively influenced by increasing age of the main farm operator. Especially the probability to choose care farming as a diversification strategy is decreasing with increasing farm operator's age. The estimated results show that broadening activities are the most negatively affected by increasing age. Consequently, the research question one: 'Are younger farmers more likely to engage in broadening activities?' can be confirmed. Overall broadening activities show to be negatively influenced by increasing age.

Amongst the different diversification activities the availability of female workforce has a negative influence on nature conservation and on- farm sale activities. For the rest of the included diversification activities, one additional woman working on the farm increases the probability to diversify. Multiple diversification activities profit the most from additional female workforce. With regards to research question two: 'Are farm households with more female workforce more likely to engage in agro-tourism and care farming activities?' the study shows a significant positive correlation of the availability of female workforce and the probability to engage in tourism and care farming activities. Thus, findings by Hassink et al. (2007), Hjalager (1996) and Nilsson (2002) are confirmed. The probability to uptake broadening activities increases relatively the most with additional female workforce.

Economic characteristics

The negative relationship of economic size and diversification probability is shown to be significant for all diversification activities included, except processing and multiple diversification activities. Processing is shown to be the only activity with a significantly increasing uptake probability with increasing SO by one unit. The results show, that deepening activities associated with agricultural production are significantly negatively influenced by increasing farm size. For activities outside traditional agriculture (broadening) the result is not significant. With increasing dependency on the share of income retrieved from the diversification activity the propensity to engage in tourism, processing, care and multiple activities increases. Contrarily, the probability to engage in sale and nature conservation decreases with increasing income dependency. These results support findings by Hjalager (1996), confirming research question four: 'Are the farms that mainly rely on income from diversification activities mainly agro- tourism farms?'. An increasing income dependency, i.e. change from one dependency category to the next (see appendix A for a detailed explanation), increases the probability to engage in tourism by around 3%. Income dependency is shown to increase the propensity to engage in broadening activities outside traditional agriculture by 14.57%.

Pasture based livestock farms and arable farms have an increased propensity to engage in nature conservation activities. Horticulture farms are more likely to diversify in activities related to the traditional side of agriculture, they increasingly engage in deepening activities, i.e. on-farm sale and processing. The probability to engage in agro- tourism and care farming are increasing for intensive livestock farms. Those activities are positively influenced when a farm holds animals. The estimated results show that research question five: 'Are farms with arable specialisation less likely to engage in on farm sale activities?' can be answered positively. Consequently the findings by Jongeneel et al. (2008) are supported by this study.

Geophysical characteristics

The population density in the farm's municipality shows to be of little significance for most diversification activities. The odds to engage in nature conservation activities decrease with increasing urbanisation. Contrarily the probability to engage in on- farm sale increases with increasing population density. This supports the theory that sale activities are favoured by farms located in more urbanized areas, gaining from lower transportation costs associated with lower distances to highly populated areas. Thus research question six: 'Are direct sale activities more likely to be situated in urban areas?', and findings by Ilbery (1991) are confirmed by the analysis. The probability to engage in broadening activities outside traditional agriculture shows to be overall negatively influenced by increasing population density.

Tourism is positively influenced by the farm being situated on dunes, the opposite holds for nature conservation. The probability to engage in on- farm sale activities increases whenever a farm is located on fluvial clays. Multiple activities are especially attractive in dune and löss hill areas. The probability to engage in nature conservation is very attractive for farms on peat soils. Thus when answering research question seven: 'Are nature conservation activities more attractive on less productive soils?' a definition of the broad concept of soil productivity is needed. Findings by Pfeifer et al. (2009) show that farmers perceive peat soils as not very fertile, thus research question seven can be answered positively.

Conclusion

In order to develop more effective policies to support farms by the choice of their risk management strategy it is important to understand the drivers leading to the risk management strategy 'diversification'. The aim of this study has been two-fold: First, we aimed to find the driving characteristics of Dutch farms favouring a diversification strategy over a specialisation strategy. Second, we aimed to explore driving factors for specific diversification activities within or outside conventional agriculture. The study analysed data from the FSS of 2011 using a binomial logit model and a multinomial probit model.

The socio-demographic characteristics included show significant results in all models estimated. The study uncovers that younger farm operators are more likely to follow a diversification strategy. This is in line with the theory proposed by Barbieri and Mahoney (2009) that young farmers need to reduce the farming risk due to the binding long term ties and larger investments attached to the farm holding. Additionally the analysis supports Mishra et al. (2004) proposing that younger farmers start off with smaller businesses and a bigger need to spread the farming risk. Once they grow older, the farm holdings grow bigger leading to less need to diversify as they can better cope with fluctuating market conditions. Available female workforce shows to positively influence the diversification decision. Moreover we find that especially care and tourism farming activities (broadening activities) benefit largely from the availability of female workforce. This supports findings by Nilsson (2002), proposing that farm women look for occupation while the men work on the fields and the theory introduced by Hjalager (1996) that those diversification activities "upgrade" ordinary household activities, making farm women's household tasks profitable.

Findings associated with the in the dataset available economic farm characteristics are more challenging to interpret. The economic size represents only the agricultural output and does not include the output derived from the diversification activity. This component can only be captured by the income dependency indicator. Care and tourism farms in particular show a significant negative diversification trend with increasing agricultural output, but a positive trend with increasing income dependency. Consequently, we may assume that farm holdings choosing very capital and labour intensive diversification activities, shift large parts of their resources to the diversification activity. This supports findings of the qualitative study from Hjalager (1996) who finds that higher investments go alongside higher income dependency. Additionally we find that increasing economic size increases the probability to uptake processing activities, showing clear evidence for economies of scale. Krugman (1990) explains that increasing returns to scale can lead to clustering patterns, or geographic concentration of an activity. This is in line with the clear clustering patterns found for on –farm processing. The theory introduced earlier, that the farm type and seasonality of the farming activity has a large influence on the diversification decision (McNally, 2001), can be supported. We find that all farming types included show larger probabilities to diversify than intensive livestock farms which are associated with constantly high labour demand and less spare time to develop a diversification strategy off season. Overall seasonal farming types favour diversification but this is not the whole picture. The study also shows that arable farms are less likely to uptake on- farm sale activities, following the theory that on farm sale is too time intensive and the hourly returns are not large enough for arable farms (Jongeneel et al., 2008). Horticulture farms on the other hand show a positive correlation with the uptake probability of on farm- sale activities, thus we can assume a higher remuneration per hour associated with on- farm sale for this farming activity. This findings underline that different farm types are associated with different remuneration per hour for on-farm sale activities, making them more or less attractive.

We find that location plays a significant role in the farms' diversification decision. Findings by Pfeifer et al. (2009) can be supported. As we showed earlier there is a cluster of diversification activities around the big national park in the centre of the Netherlands, confirming that the attractiveness of the landscape can lead to clustering of complementary diversification activities due to spill over effects. Additionally we showed that recreational activities benefit from the higher attractiveness of landscapes associated with lower population density. The study can confirm the theory introduced by Ilbery (1991), that diversification activities benefit from proximity to urban areas only partially. On- farm sale activities have been found to benefit from higher population density. This activity is associated with shortening and improving the quality of the food supply chain which is more attractive for farms closer to the demand side of the food market. Better direct marketing options are assumed to lead to higher turnovers and reduced the risk of price volatility. For the general diversification decision this cannot be supported by this analysis. Consequently here finings by Lange et al. (2013) based on land pricing theory, support the negative correlation of population density and diversification probability. They conclude that higher proximity to urban areas leads to higher opportunity costs of land and labour making an exit strategy more attractive with increasing farming risk. With respect to the soil type and its productivity we find that nature conservation activities are especially attractive for farms on peat soils. One explanation can be found in survey results from Pfeifer et al. (2009). They reveal that farmers describe peaty soils as less productive, leading to an increasing need to look for additional income outside traditional agriculture. Another explanation can be found in the content of the nature conservation support programs in the Netherlands. Large parts of the nature conservation measures in the Netherlands are aiming at meadow bird protection. Beintema et al. (1995, p. 325) show, that peat soils are the preferred habitat for meadow birds, making nature conservation especially attractive for farm holdings specialized in pasture based livestock or arable farming on peat soils.

When interpreting diversification as a risk management strategy, the results show that this strategy is chosen over a specialisation or a 'business as usual' strategy by farms with younger main farm operators, with more female workforce available and smaller average economic output. Broadening activities are often chosen by, small farms (in terms of agricultural output), which have allocated large parts of the farms resources away from the main farming activity to the diversification activity, in order to ensure the farms' survival. Deepening activities, associated with the agricultural production side, are less positively influenced by the higher income dependency, thus those activities may be motivated by the farms' wish for a stable additional income. The study finds significant differences in the drivers leading to a combination of broadening and deepening and the drivers leading to the farms' focus on one of the categories. This supports the distinction made in this study and emphasises the importance to investigate further in the drivers leading to the uptake of a combination of diversification activities.

The analysis was able to confirm several findings from other researchers with the newest data available for the Dutch agricultural sector but also to provide some insights in more specific diversification strategies. The study is based on farm characteristics included in the FSS, which makes the results comparable with other studies over time and space, but also generates some limitations. Additional variables capturing psychological aspects could give further insights into the farmers' attitude towards risk. This, more qualitative approach, is one suggestion for further research. In line with this, a suggestion is a behavioral economic analysis, one could think of a game simulation with different resource allocations. Moreover future research could attempt to use the income dependency indicator as weighting measure for each diversification activity, leading to a scaling of the dependent variable. Furthermore, the data set does not include information on the income gained from off-farm labour, which could also lead to a better understanding of the farms' resource allocation. Additional research exploring the influence of spatial and geographical aspect on the choice of the farming activity could give this research a supplementary dimension.

Concluding, we find by comparing the result of the general model with the two specific diversification models, that in order to analyse diversification characteristics adequately, a thorough analysis of the different diversification activities is necessary. Although, most of the specific diversification activities are driven by the same factors as the general diversification decision, it is necessary to account for those differences when making policies aiming to support specific activities. As shown in the previous section, the results are very dependent on the classification of diversification activities. Thus, when comparing the growing literature on diversification motives it is necessary to take a close look at the still very ambiguous definitions of activities and their categorisation.

Acknowledgments

Especially I thank the Dutch Agriculture Economics Research Institute (LEI) for making the survey data available. I am grateful to the supervisors Prof. Dr. W. Heijman, Dr. T. Kuhlman and Dr. C. Wieck for their support and inspiration throughout the process. Furthermore, I thank Esther Boere for her detailed and constructive comments.

References

Aguglia, L., Henke, R., Poppe, K., Roest, A., Salvioni, C., 2009. Diversification and multifunctionality in Italy and the Netherlands: a comparative analysis, Contributions Second meeting Wye City Group Conference. FAO, Rome.

Baldwin, R.E., Wyplosz, C., 2004. The economics of European integration. McGraw-Hill London.

Barbieri, C., Mahoney, E., 2009. Why is diversification an attractive farm adjustment strategy? Insights from Texas farmers and ranchers. Journal of Rural Studies 25, 58-66.

Beintema, A.J., Moedt, O., Ellinger, D., 1995. Ecologische Atlas van de Nederlandse weidevogels. Schuyt & Co., Haarlem.

Berkhout, P., Roza, P., 2012. Agricultural Economic Report 2012 of the Netherlands: summary. LEI, part of Wageningen UR, The Hague.

Communication department of the European Commission, Netherlands. http://europa.eu/about-eu/countries/member-countries/netherlands/. Retrieved 13.06.2013.

Daganzo, C., 1979. Multinomial Probit: The Theory and Its Application to Demand Forecasting. Academic Press, New York.

Dernoi, L.A., 1983. Farm tourism in Europe. Tourism Management 4, 155-166.

Di Iacovo, F., Vadnal, K., O'Connor, D., 2009. Supporting policies for social farming in Europe: progressing multifunctionality in responsive rural areas. Arsia-Agenzia Regionale per lo Sviluppo e l'Innovazione nel settore Agricolo-forestale.

European Commission, 2008. Commission Regulation (EC) No 1242/2008 of 8 December 2008 establishing a Community typology for agricultural holdings, L 335/3. Official Journal of the European Union.

Finocchio, R., Esposti, R., 2008. Determinants of farm diversification and interaction with the CAP. An application to FADN of Marche region (Italy), 12th Congress of the European Association of Agricultural Economists–EAAE.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., 2010. Multivariate data analysis. Pearson, Upper Saddle River, NJ.

Hardy, M.A., Bryman, A., 2004. Handbook of data analysis. SAGE Publications Limited.

Hassink, J., Zwartbol, C., Elings, M., Thissen, J., 2007. Current status and potential of care farms in the Netherlands. NJAS-Wageningen Journal of Life Sciences 55, 21-36.

Heringa, P.W., Van der Heide, M.M., Heijman, W.J., 2012. The economic impact of multifunctional agriculture in The Netherlands, 2012 Conference, August 18-24, 2012, Foz do Iguacu, Brazil. International Association of Agricultural Economists.

Hjalager, A.-M., 1996. Agricultural diversification into tourism: Evidence of a European community development programme. Tourism management 17, 103-111.

Ilbery, B.W., 1991. Farm diversification as an adjustment strategy on the urban fringe of the West Midlands. Journal of Rural Studies 7, 207-218.

Jongeneel, R.A., Polman, N.B.P., Slangen, L.H.G., 2008. Why are Dutch farmers going multifunctional? Land Use Policy 25, 81-94.

Kleijn, D., Berendse, F., Smit, R., Gilissen, N., 2001. Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. Nature 413, 723-725.

Knickel, K., Renting, H., 2000. Methodological and Conceptual Issues in the Study of Multifunctionality and Rural Development. Sociologia Ruralis 40, 512-528.

Kostov, P., Lingard, J., 2003. Risk management: a general framework for rural development. Journal of Rural Studies 19, 463-476.

Kropko, J., 2007. Choosing between multinomial logit and multinomial probit models for analysis of unordered choice data. ProQuest.

Krugman, P., 1990. Increasing returns and economic geography. National Bureau of Economic Research.

Lange, A., Piorr, A., Siebert, R., Zasada, I., 2013. Spatial differentiation of farm diversification: How rural attractiveness and vicinity to cities determine farm households' response to the CAP. Land Use Policy 31, 136-144.

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior.

McNally, S., 2001. Farm diversification in England and Wales—what can we learn from the farm business survey? Journal of Rural Studies 17, 247-257.

Mishra, A.K., El-Osta, H.S., Sandretto, C.L., 2004. Factors affecting farm enterprise diversification. Agricultural finance review 64, 151-166.

Nilsson, P.Å., 2002. Staying on farms: An ideological background. Annals of tourism research 29, 7-24.

OECD, 2001. Multifunctionality: towards an analytical framework. OECD Publishing.

Pfeifer, C., Jongeneel, R.A., Sonneveld, M.P.W., Stoorvogel, J.J., 2009. Landscape properties as drivers for farm diversification: A Dutch case study. Land Use Policy 26, 1106-1115.

Pope, R.D., Prescott, R., 1980. Diversification in relation to farm size and other socioeconomic characteristics. American Journal of Agricultural Economics 62, 554-559.

Robison, L.J., Barry, P.J., 1987. The competitive firm's response to risk. Macmillan New York.

Roest, A.E., Oltmer, K., Venema, G.S., Jager, J.H., Schoorlemmer, H.B., Hendriks-Goossens, V.J.C., Dekking, A.J.G., Kamstra, J.H., Visser, A.J., Jeurissen, L.J.J., Gies, T.J.A., 2009. Kijk op multifunctionele landbouw, omzet en impact : achtergronddocument. LEI Wageningen UR, Den Haag, p. 141.

Van Der Ploeg, J.D., Renting, H., Brunori, G., Knickel, K., Mannion, J., Marsden, T., De Roest, K., Sevilla-Guzmán, E., Ventura, F., 2000. Rural Development: From Practices and Policies towards Theory. Sociologia Ruralis 40, 391-408.

Van der Ploeg, J.D., Roep, D., 2003. Multifunctionality and rural development: the actual situation in Europe, in: Durand, G.v.H.a.G. (Ed.), Multifunctional Agriculture. A New Paradigm for European Agriculture and Rural Development. Ashgate, Aldershot, Hampshire, England, pp. 37-54.

Van Huylenbroeck, G., Vandermeulen, V., Mettepenningen, E., Verspecht, A., 2007. Multifunctionality of agriculture: a review of definitions, evidence and instruments. Living Reviews in Landscape Research 1, 1-38.

Verbeek, M., 2008. A guide to modern econometrics. Wiley.

Vik, J., McElwee, G., 2011. Diversification and the Entrepreneurial Motivations of Farmers in Norway. Journal of Small Business Management 49, 390-410.

Vries, W.M.D., 1993. Farming with other gainful activities in the Netherlands. Sociologia Ruralis 33, 190-202.

Wilson, G.A., 2007. Multifunctional agriculture a transition theory perspective. CABI, Wallingford.

Appendix A

The structure of the Netherlands

The kingdom of the Netherlands is a constitutional monarchy in Western Europe bordering the North Sea, Belgium and Germany. In addition to the European area six Caribbean islands also belong to the Netherlands. The Netherlands are with 479 inhabitants per km^2 the most densely populated country in the European Union. It is a relatively small country covering a total of 41,526 km^2, of which 34,564 km^2 are land the rest is water area. Large parts of the Netherlands are below sea level (ca. 24%) and about half of the land is less than one meter above sea level. Only in the south-east of the country are some hilly areas. Administratively the country is divided into twelve provinces: Drenthe, Flevoland, Friesland (Fryslân), Gelderland, Groningen, Limburg, North Brabant (Noord-Brabant), North Holland (Noord-Holland), Overijssel, Utrecht, Zeeland and South Holland (Zuid-Holland). These provinces are further divided into 408 municipalities (gemeenten) (as of January 2013). Dutch agriculture benefits from a positive combination of geographical and natural conditions leading to an intensive agricultural sector. These are amongst others: its mild climate, its good trading position with big ports and rivers as well as its flat landscape and fertile soils (Communication department of the European Commission, 13.06.2013)

Diversification on provincial level

The province in which a farm is located is also potentially highly valuable to detect diversification patterns. Table A.1 depicts the distribution of activities on provincial level. The largest share of diversifying farms can be found in Friesland and Gelderland (15%). Flevoland has the fewest diversified farms. Tourism and care farms as well as mixed farms and other activities are mostly situated in Gelderland. On- farm sale is most common in North- Brabant. Nature conservation is most common in Friesland and on- farm processing is most common in North- Holland.

Table A. 1 Distribution of diversification activities on province level in per cent of total diversification

	Tourism	Sale	Nature conservation	Processing	Care	Mixed activities	Other activities	Diversifying farms
Drenthe	7%	4%	3%	4%	7%	4%	3%	4%
Flevoland	1%	2%	1%	12%	1%	2%	3%	2%
Friesland	5%	3%	24%	6%	4%	10%	9%	15%
Gelderland	19%	17%	14%	7%	20%	16%	18%	15%
Groningen	2%	4%	9%	6%	4%	4%	6%	7%
Limburg	7%	14%	5%	4%	7%	9%	4%	7%
North-Brabant	16%	23%	5%	10%	20%	11%	14%	11%
North-Holland	7%	7%	8%	17%	9%	11%	10%	9%
Overijssel	13%	7%	8%	7%	13%	9%	11%	9%
Utrecht	4%	3%	8%	3%	4%	7%	8%	7%
South-Holland	6%	9%	11%	16%	8%	11%	11%	10%
Zeeland	13%	8%	3%	8%	4%	6%	3%	6%

Agricultural structure

In the year 2011, 55% of the total Dutch land area is used as agricultural area; this represents an area of 1,858 million hectares. This includes all land that is permanently or temporarily part of a farm holding, and is mainly used for the production of agricultural goods (crops, livestock), including fallow land and (temporary) grassland. The agricultural utilised land is cultivated by 70,394 farm holdings, employing a total workforce of 209,000 people. On average the utilised agricultural area is 34.3 ha per farm holding. The yearly standard output is on average €37,8000. Most of the agricultural holdings in the Netherlands are family farms; this means that the labour and management of the farm is carried out by family members (Berkhout and Roza, 2012).

In 2011, the highest production on arable land yield fodder maize with a total of 10,559 million kg. In horticultural production tomatoes represent the largest share with 815 million kg. In 2011 a livestock of 9 million grazing animals and 222 million non-grazing animals are held. The primary Dutch agriculture and horticulture sector's gross production value lies at 25.5 billion Euros in 2011. This represents a share of 1.36% of the Gross Domestic Product (Berkhout and Roza, 2012).

Multiple diversification activities

It is a common problem within data analysis of diversification activities that farm holdings engage in several diversification activities simultaneously. Unfortunately the available data does not allow a distinction between the different activities by main source of income. For the second analysis these 2,238 cases of multiple diversification are grouped together, this is in line with other researchers (Finocchio and Esposti (2008) and Vik and McElwee (2011)). Amongst all diversifying farms only 18% do so by multiple activities. Figure A.1 gives an overview of the distribution of multiple and single diversification activities within each diversification activity.

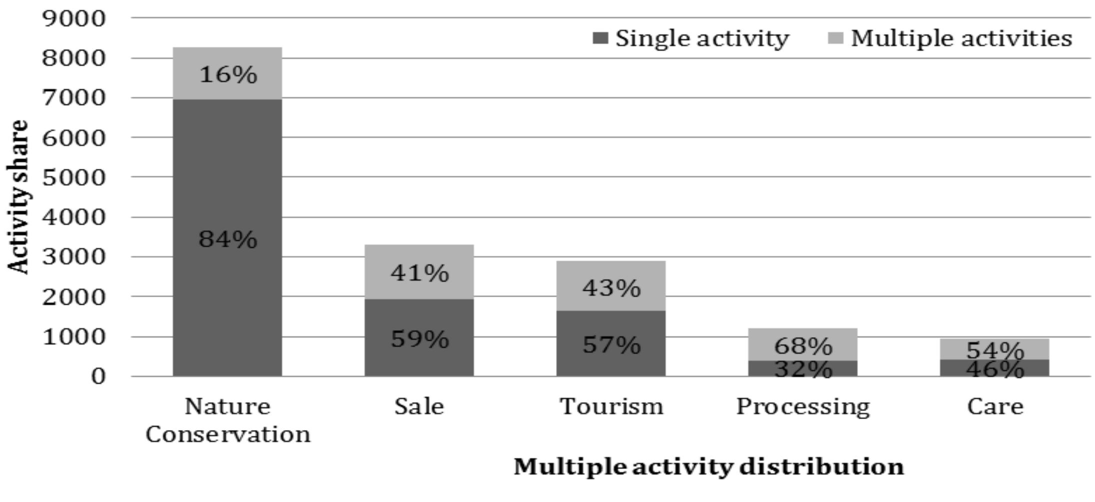

Figure A. 1 Distribution of multiple diversification activities in percentage of total number of diversifiers

The largest share of diversifying farms, a total of 6,815 farms engage in nature conservation activities. 84% of them precede no other additional diversification activity, only 16% of them carry out multiple activities. The data set includes 1,566 farms which engage in agro- tourism activities. 26% of them do so in combination with one or more other diversification activities. A total of 1,909 farms, carry out on- farm sale activities of which 34% engage in multiple activities.

On- farm processing of farm products is pursued by a total of 390 farms this activity is most commonly combined with another activity. Processing is mainly combined with direct sale. Care farms have the smallest share in the sample with a total of 393 farms pursuing care activities. 38% of them engage in multiple diversification activities, mainly in combination with nature conservation. The most popular combination is nature conservation and tourism, carried out by 23% of the sum of farms engaging in multiple activities. Followed by the combination of on farm processing and sale with a share of 14% of all multiple diversifying farms. A combination of tourism and sale is carried out by 9% of all farms engaging in several diversification activities.

<u>Income dependency indicator</u>

The income dependency indicator reflects the percentage share of the overall farm's income derived from the diversification activity. Consequently this variable is applicable to all farms that do not diversify, and thus it is not included in the first analysis. The income dependency is indicated on a scale from 1 to 4 where 1 = <10%, 2 = 10-30%, 3 = 30-50% and 4 = > 50%. Due to the fact that the categories are not identically large the analysis assumes a normal distribution of the real income dependency in category 3 and 4, leading to the estimated mean values 2 = 20% and 3 = 40%. For category 4 it is reasonable to assume that there is only a very small share of farms with an income dependency above 90% and for certain none with 100%; as then the holding would not appear in the FSS. Consequently one can assume a positively skewed distribution in this category leading to an expected mean of 65%. The reverse is true for the first category where smaller values then 1% are very unlikely and the distribution is assumed to be negatively skewed, to keep in line with category 4 we assume an expected mean value of 6.5%. This variable is only relevant for the second analysis as for the first analysis the income dependency indicator is perfectly correlated with the explanatory variable. The distribution of income dependencies amongst the different diversification activities can be seen in Figure A. 2 A.2.

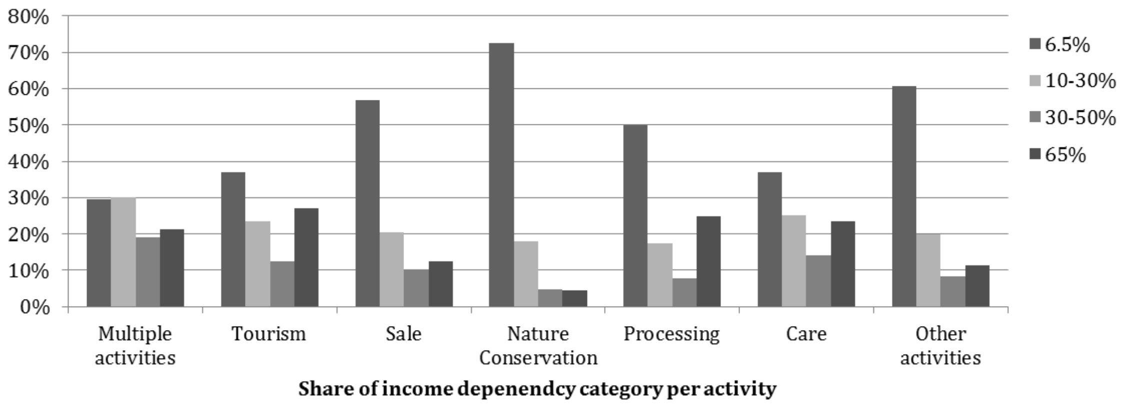

Figure A. 2 Income dependency distribution on different diversification activities

Appendix B

Descriptive Statistics

Table B. 1 Summary statistic for explanatory variables binomial diversification model

	All farms N = 68724		No diversification N = 54830		Diversification N = 13894	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Age	54.317	11.740	54.725	11.975	52.706	10.615
Fem	1.048	2.164	1.019	2.315	1.162	1.415
EconSize	278596.1	622885.3	295774.5	678880.6	210804.8	307539.1
IncDep	0.468	0.944	0.140	0.555	1.761	1.048
Farm type						
Past	0.524	0.499	0.496	0.500	0.632	0.482
Arab	0.169	0.375	0.176	0.381	0.144	0.351
Hort	0.168	0.374	0.182	0.386	0.116	0.321
Inte	0.087	0.282	0.098	0.298	0.042	0.201
Mix	0.051	0.221	0.048	0.213	0.066	0.249
PopDens	390.528	920.513	391.893	908.691	385.140	965.769
Soil type						
Mc	0.244	0.429	0.235	0.424	0.280	0.449
Ps	0.533	0.499	0.564	0.496	0.412	0.492
Pe	0.063	0.243	0.049	0.217	0.116	0.320
Fc	0.123	0.328	0.120	0.325	0.134	0.341
Du	0.022	0.147	0.022	0.145	0.024	0.154
Lh	0.015	0.123	0.011	0.103	0.034	0.181

Table B. 2 Summary statistics for explained variables

All farms N=68724	Mean	Std Dev	Skewdness	Kurtosis
Diversification	0.20	0.40	1.48	3.20
Multiple diversification	0.03	0.18	5.27	28.74
Tourism	0.02	0.15	6.25	40.06
Sale	0.03	0.17	5.69	33.34
Nature Conservation	0.10	0.30	2.64	7.98
Processing	0.01	0.08	13.11	172.88
Care	0.01	0.08	12.55	158.58
Other activities	0.01	0.09	10.72	115.89
Broadening and Deepening	0.02	0.13	7.38	55.53
Broadening	0.15	0.35	1.99	4.98
Deepening	0.04	0.19	4.84	24.44

Table B. 3 Summary statistic for explanatory variables binomial specific activity system

	Multiple diversification N = 2531		Tourism N = 1566		Sale N = 1909		Nature conservation N = 6815		Processing N = 390		Care N = 393		Other activities N = 290	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Age	51.52	9.81	53.98	10.68	52.10	10.18	53.39	11.06	50.64	10.18	50.26	8.81	50.26	9.53
Fem	1.45	1.56	1.34	1.47	1.40	2.00	0.90	0.65	1.64	3.44	1.51	2.19	1.18	0.96
EconSize	202385.6	259131.7	114468.2	207350.0	254827.9	468085.9	210284.1	232998.1	440828.7	739119.5	179871.5	209248.3	259520.2	256898.7
IncDep	2.24	1.12	2.32	1.22	1.78	1.06	1.41	0.78	2.08	1.26	2.25	1.17	1.49	0.92
Farm type														
Past	0.61	0.49	0.65	0.48	0.16	0.37	0.79	0.41	0.22	0.42	0.66	0.47	0.67	0.47
Arab	0.11	0.31	0.17	0.38	0.19	0.39	0.13	0.34	0.34	0.48	0.05	0.23	0.11	0.32
Hort	0.15	0.35	0.07	0.25	0.44	0.50	0.02	0.13	0.32	0.47	0.10	0.30	0.05	0.22
Inte	0.04	0.20	0.05	0.21	0.08	0.27	0.03	0.16	0.03	0.18	0.10	0.31	0.09	0.29
Mix	0.10	0.29	0.06	0.23	0.12	0.33	0.04	0.20	0.08	0.27	0.08	0.28	0.07	0.26
PopDens	417.10	535.19	395.71	483.75	468.13	1685.63	347.29	952.83	378.38	447.58	363.71	388.79	430.50	576.46
Soil type														
Mc	0.26	0.44	0.25	0.43	0.28	0.45	0.28	0.45	0.55	0.50	0.19	0.39	0.24	0.43
Ps	0.41	0.49	0.57	0.49	0.50	0.50	0.35	0.48	0.25	0.43	0.61	0.49	0.54	0.50
Pe	0.11	0.32	0.03	0.16	0.02	0.14	0.17	0.38	0.06	0.24	0.06	0.24	0.06	0.24
Fc	0.13	0.34	0.08	0.27	0.16	0.36	0.14	0.35	0.09	0.29	0.12	0.32	0.15	0.36
Du	0.04	0.20	0.04	0.20	0.02	0.13	0.02	0.13	0.04	0.20	0.01	0.07	0.01	0.08
Lh	0.05	0.21	0.03	0.16	0.04	0.18	0.03	0.18	0.01	0.11	0.02	0.13	0.00	0.00

Table B. 4 Summary statistics for classification of activities

	Broadening and Deepening N = 1195		Broadening N = 10096		Deepening N = 2603	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Age	51.49	9.92	53.09	10.78	51.76	10.19
Fem	1.50	1.69	1.04	1.02	1.48	2.27
EconSize	199582.40	253338.50	193395.60	231899.00	283479.80	510284.00
IncDep	2.35	1.09	1.66	1.00	1.89	1.12
Farm type						
Past	0.49	0.50	0.76	0.43	0.19	0.39
Arab	0.11	0.32	0.13	0.34	0.21	0.41
Hort	0.23	0.42	0.03	0.17	0.41	0.49
Inte	0.04	0.20	0.03	0.18	0.07	0.26
Mix	0.13	0.34	0.05	0.21	0.11	0.32
PopDens	451.51	573.83	362.27	831.76	443.37	1462.00
Soil type						
Mc	0.25	0.43	0.27	0.44	0.33	0.47
Ps	0.42	0.49	0.40	0.49	0.45	0.50
Pe	0.09	0.28	0.14	0.35	0.03	0.17
Fc	0.14	0.35	0.13	0.34	0.14	0.35
Du	0.04	0.19	0.02	0.15	0.02	0.15
Lh	0.05	0.23	0.03	0.18	0.03	0.17

Appendix C

The binary choice model

For dichotomous dependent variables, binary choice models estimate the probability of the outcome to be one. In applied work binomial logit and binomial probit model are most commonly used (Verbeek, 2008, p. 201). Binary choice models can be derived from an underlying utility maximisation problem leading to a latent variable representation of the model. The utility difference between two alternative choices is furthermore denoted as y_i^*, and assumed to be dependent on a vector of individual observed characteristics (x_i') and unobserved characteristics (ε_i):

$$y_i^* = \beta_i x_i^{'} + \varepsilon_i.$$

If the utility difference between two choices is greater than zero $\mathrm{y_i^*} > 0$, one can observe a positive outcome (i.e. $\mathrm{y_i} = 1$)). Consequently, the probability that the observable outcome is one ($\mathrm{P(y_i = 1 \mid x_i)}$) is equal to the probability that the latent variable is positive ($\mathrm{P(y_i^* > 0 \mid x_i')}$). Probit and logit models differ in the distribution function of the error term ε_i. Where the probit model assumes a standard normal distribution function the logit model assumes the standard logistic distribution function. The logit model is more common, and yields only marginally different results (Verbeek, 2008, p. 203). To improve accuracy when comparing this study with other research results, this analysis uses logit models.

Furthermore it can be shown that if the error terms are assumed to be standard logistic distributed the following holds:

$$P(y_i = 1|x_i) = P(y_i^* > 0|x_i') = P(y_i^* > 0|x_i) = P(\beta_i x_i + \varepsilon_i > 0) = P(\varepsilon_i < \beta_i x_i) = logit^{-1}(\beta_i x_i) = \frac{1}{1+e^{-\beta_i x_i}} = p_i \, .$$

The logistic regression model is generally defined as follows (Verbeek, 2008, p. 203):

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_i x_i.$$

The left hand side of the equation is referred to as the log odds ratio. Consequently the coefficients of the estimated general diversification model displayed in table C.1 can be interpreted as describing the effects upon the odds ratio. The parameters in the analysis are estimated using maximum likelihood method.

Table C. 1 Compared binary choice models for the diversification model

	Logit				**Probit**			
Variable	Coef.	Std. Err.	z	P>z	Coef.	Std. Err.	z	P>z
AGE	-0.020	0.001	-23.330	0.000	-0.012	0.000	-23.270	0.000
FEM	0.076	0.006	13.390	0.000	0.035	0.002	14.410	0.000
ECONSIZE	-0.042	0.003	-13.130	0.000	-0.022	0.002	-13.640	0.000
PAST	0.796	0.047	16.800	0.000	0.439	0.025	17.790	0.000
ARAB	0.289	0.053	5.410	0.000	0.157	0.028	5.590	0.000
HORT	-0.052	0.054	-0.960	0.338	-0.013	0.028	-0.470	0.640
MIX	0.928	0.060	15.510	0.000	0.518	0.033	15.800	0.000
POPDENS	-0.005	0.002	-2.640	0.008	-0.002	0.001	-2.630	0.009
MC	-0.539	0.038	-14.110	0.000	-0.323	0.023	-13.930	0.000
PS	-1.144	0.036	-31.860	0.000	-0.655	0.022	-30.250	0.000
FC	-0.690	0.042	-16.460	0.000	-0.404	0.025	-16.030	0.000
DU	-0.333	0.072	-4.610	0.000	-0.234	0.043	-5.490	0.000
LH	0.437	0.072	6.080	0.000	0.267	0.044	6.040	0.000
Log-likelihood				-32706.509				-32746.83
LR χ^2 (13)				3777.95				3697.30
McFadden's R^2				0.0546				0.0534

Goodness of fit

The overall goodness-of-fit in binary models can be tested in various ways, some of them are reported in table C.2. The Log-likelihood function is always negative and approaches zero with a perfect fit, in this case the model estimates a Log- likelihood value of -32706.509. The large value is to be explained by the large amount of data analysed. The likelihood ratio chi-square (LR χ^2) and can be seen as overall measurement of the model significance. It tests the null hypothesis that all coefficients are zero except the constant. The test shows for the general model a significant LR χ^2 value of 3,777.95; reflecting an improvement of fit moving from the null model to the model including all parameters. The 13 degrees of freedom represent the 13 variables included in the general diversification model. The McFadden R^2 compares two likelihood values, logLFull measuring the maximum value of the model including all independent variables, and logLIntercept measuring the maximum of the likelihood function when all parameters except the intercept are set to zero. Unlike the R^2 used in linear regression models the McFadden R^2 is not measured in terms of variance, since the variance in logistic regression is fixed at the variance of the standard logistic distribution. It expresses the deviance of the model as a proportion of deviance for the null model (Verbeek, 2008, p. 205). The McFadden R^2 reveals that the model performs 5.46% better than a model that specifies the probability of diversifying to be constant.

Table C. 2 Different goodness-of-fit measures logit model

Log-Lik Intercept Only:	-34595.482	Log-Lik Full Model:	-32706.509
D(68710):	65413.018	LR(13):	3777.945
		Prob > LR:	0.000
McFadden's R2:	0.055	McFadden's Adj R2:	0.054
ML (Cox-Snell) R2:	0.053	Cragg-Uhler (Nagelkerke) R2:	0.084
McKelvey & Zavoina's R2:	0.108	Efron's R2:	0.061
Variance of y*:	3.687	Variance of error:	3.29
Count R2:	0.798	Adj Count R2:	0.002
AIC:	0.952	AIC*n:	65441.018
BIC:	-699868.914	BIC':	-3633.153
BIC used by Stata:	65568.948	AIC used by Stata:	65441.018

An alternative way of measuring the goodness-of-fit is the comparison of correct and incorrect predictions (see table C.3). The table shows a one if the estimated probability is larger than 0.5 and zero otherwise. In the group of diversifying farms there are 48% of cases in which the model predicts a high chance of diversification but they do not. In the group of not diversifying farms around 20% cases are classified wrong. The model yields a sum of the proportions of correct predictions of 79.82%. In other words, around 80% of the estimated probability outcomes are correctly predicted in the model. One has to note that the cross-tabulation of actual and predicted outcomes does not reflect how large the deviation of the predicted outcomes is. In addition to that it holds that the outcome is highly influenced by threshold of 0.5. Consequently the interpretation is not as meaningful as the other goodness-of-fit measures (Verbeek, 2008, p. 207).

Table C. 3 Cross-tabulation of actual and predicted outcome logit model

Predicted probabilities	**True probabilities**		
	Diversifying	Not diversifying	
Diversifying	51.93%	48.07%	
Not diversifying	19.92%	80.08%	
Correctly classified			**79.82%**

Significance of variables

In order to test the significance of the individual parameters LR χ^2 test was used. This test compares the likelihood ratio of the full model including all parameters with the likelihood ratio of a constrained model excluding one parameter at the time (Verbeek, 2008, p. 182). A list including the results for all variables can be found in table C.4. Leading to the conclusion that we cannot reject H0 for the dummy variable 'hort', i.e. the effects of this variable is insignificant for the general model. A non-significant likelihood ratio test implies no difference between the full model and the model without the given variable.

Table C. 4 Likelihood ratio test for logit regression

diversification	Df	Chi2	P>Chi2	-2*log ll	Res. Df	AIC
Original Model				65413.02	68710	65441.02
-age	1	556.72	0.0000	65969.74	68709	65995.74
-fem	1	205.23	0.0000	65618.25	68709	65644.25
-econsize	1	230.45	0.0000	65643.47	68709	65669.47
-past	1	324.56	0.0000	65737.58	68709	65763.58
-arab	1	30.13	0.0000	65443.14	68709	65469.14
-hort	1	0.91	0.3394	65413.93	68709	65439.93
-mix	1	246.82	0.0000	65659.83	68709	65685.83
-popdens	1	10.96	0.0009	65423.98	68709	65449.98
-mc	1	194.99	0.0000	65608.01	68709	65634.01
-ps	1	945.35	0.0000	66358.37	68709	66384.37
-fc	1	269.39	0.0000	65682.41	68709	65708.41
-du	1	21.89	0.0000	65434.91	68709	65460.91
-lh	1	36.62	0.0000	65449.64	68709	65475.64

Terms dropped one at a time in turn.

Multicollinearity

Testing for multicollinearity, shows no significant approximate relationship amongst the regressions' parameters. In other words, the independent variables can be assumed to be no linear predictions of each other. Verbeek (2008, p. 43) recommends that variance inflation factors (VIF) can be used to detect multicollinearity. VIF indicate the factor by which the variance of the standard error is inflated compared to the situation where there is no correlation between the variable and any of the other explanatory variables. Furthermore it is suggested that a VIF lower than 10 indicates no signs of multicollinearity, this is accurate for all explanatory variables included, as displayed in table C.5. Additionally Verbeek (2008, p. 44) proposes that R^2 values close to one indicate that an independent variable can be closely approximated by a linear combination of the other regressors. However, the sample is sufficiently large and the variance of the error term is sufficiently small (see table C.1) so that the large R^2 values do not need to cause problems. Another way to test for multicollinearity is to analyse the variance- covariance matrix (see table C.6). The matrix additionally shows no signs of collinearity amongst the independent variables.

Table C. 5 Collinearity diagnostics general logit model

Collinearity Diagnostics

Variable	VIF	SQRT VIF	Tolerance	R-Squared
age	1.05	1.02	0.9558	0.0442
fem	1.14	1.07	0.8757	0.1243
econsize	1.27	1.13	0.7896	0.2104
past	3.61	1.90	0.2771	0.7229
arab	2.75	1.66	0.3630	0.6370
hort	2.69	1.64	0.3716	0.6284
mix	1.55	1.25	0.6442	0.3558
popdens	1.02	1.01	0.9787	0.0213
mc	3.85	1.96	0.2597	0.7403
ps	4.55	2.13	0.2199	0.7801
fc	2.61	1.61	0.3838	0.6162
du	1.36	1.17	0.7351	0.2649
lh	1.24	1.11	0.8054	0.1946
Mean VIF	2.21			

	Eigenval	Cond Index
1	4.4987	1.0000
2	1.4246	1.7770
3	1.2108	1.9276
4	1.0488	2.0711
5	1.0194	2.1008
6	0.9998	2.1212
7	0.9463	2.1804
8	0.8104	2.3561
9	0.7370	2.4707
10	0.6255	2.6819
11	0.5404	2.8854
12	0.0754	7.7225
13	0.0487	9.6147
14	0.0145	17.6031

Condition Number 17.6031

Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)

Det(correlation matrix) 0.0355

Table C. 6 Covariance matrix for general logit model

Covariance matrix of coefficients of logit model

e(V)	age	fem	econsize	past	arab
age	7.590e-07				
fem	-1.260e-07	.00003247			
econsize	4.773e-07	-6.561e-06	.00001033		
past	-1.338e-06	-.00002108	.00003442	.00224612	
arab	-1.981e-06	-.00001777	.00003949	.00212507	.00285542
hort	-1.161e-07	-.00004469	.00001266	.00203776	.00207082
mix	-1.525e-06	-.00001908	.00002608	.00205492	.0021146
popdens	-6.046e-08	-3.900e-07	7.298e-08	-7.494e-08	2.880e-06
mc	8.410e-07	-1.782e-06	-6.983e-06	.0000197	-.000214
ps	5.484e-07	-3.733e-06	-3.022e-07	.00016193	.00009123
fc	3.438e-07	-2.883e-06	1.936e-06	.00008753	.00003013
du	-4.128e-07	.00001165	-2.260e-06	.00002656	-.00007384
lh	-1.242e-06	-5.606e-07	4.578e-06	.00008653	-.00018325
_cons	-.00004031	.00001299	-.00006952	-.002151	-.00207631

e(V)	hort	mix	popdens	mc	ps
hort	.00291577				
mix	.00202902	.00357709			
popdens	-8.367e-06	1.269e-06	4.318e-06		
mc	-9.269e-06	-.0001218	3.017e-06	.00145927	
ps	.0002203	.00005548	7.312e-06	.00106562	.00128884
fc	.00007076	-.00001284	2.612e-06	.00105467	.0010525
du	-.00035049	-.00007619	-6.050e-06	.001055	.00099783
lh	.00004426	-.00012705	-8.561e-06	.00110156	.00105061
_cons	-.00209426	-.00199872	-.00001654	-.00108362	-.0012602

e(V)	fc	du	lh	_cons
fc	.00175725			
du	.00103415	.00520876		
lh	.00105398	.00107213	.00515903	
_cons	-.00114514	-.00094571	-.00099896	.0055215

The multinomial choice model

To estimate probabilities amongst different unordered categorical dependent variables, probabilistic choice models can be used. These model each choice as separate equation including the predictors and the error. The advantage compared to a series of binary choice models is that multinomial choice models use for all outcomes the same sample size leading to more accurate predicted probabilities. Multinomial logit (MNL) and multinomial probit (MNP) models are most widely used (Hardy and Bryman, 2004, p. 277). When choosing between multinomial logit and multinomial probit models Kropko (2007, p. 1) points out that it is much of a choice between accuracy of the multinomial probit model and computational ease of the multinomial logit model. Multinomial logit models can be seen as an extension of binary logit models. The MNL model estimates simultaneously binary logits for all comparisons among the outcome categories. The biggest difference of the described MNL model and the MNP model is the assumption of independence of irrelevant alternatives (IIA) that has to hold for the MNL model. IIA requires that the individual's evaluation of an alternative relative to another alternative (i.e. the probabilities to choose an alternative) does not change if a third alternative is added to or dropped from the analysis. That assumption is strongly related to the biggest difference of the two models: the distribution of the error terms. In the MNL model the error terms are assumed to be independent, consequently forcing the IIA to hold. The MNP model does not assume uncorrelated error terms, and is consequently assumed to perform better (Kropko, 2007, p. 2).

When running the MNL model for our data the IIA test according to the Hausman and Small-Hausman indicate a violation of the IIA. Both results are presented in table C.7. Hardy and Bryman (2004, p. 281) confirm that the tests deliver in practice often inconsistent results, and consequently they recommend to follow McFadden (1973), suggesting that IIA implies a use of MNL only when the outcome categories cannot possibly assumed to be substitutes. The analysed data set and choices categorized for this analysis cannot support this assumption, broadening, deepening and combinations of the two can be assumed to be substitutable. It is for example very realistic to assume that when adding the choice to engage in broadening and deepening activities simultaneously the probabilities to engage in one of the other two (only broadening and only deepening) will change. Furthermore, Kropko (2007, p. 2) points out that with a violation of the IIA the MNL is an incorrect specified model, leading to inconsistent MNL coefficient estimates.

Table C. 7 IIA test for multinomial logit

**** Hausman tests of IIA assumption (N=68724)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

Omitted	chi2	df	P>chi2	evidence
1	274.387	30	0.000	against Ho
2	-4.4e+03	30	---	---
3	4357.587	30	0.000	against Ho

Note: If chi2<0, the estimated model does not meet asymptotic assumptions of the test.

**** Small-Hsiao tests of IIA assumption (N=68724)

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

Omitted	lnL(full)	lnL(omit)	chi2	df	P>chi2	evidence
1	-1.12e+04	-1.11e+04	34.123	30	0.276	for Ho
2	-5598.480	-5063.841	1069.278	30	0.000	against Ho
3	-9705.325	-9658.045	94.559	30	0.000	against Ho

In the MNP model the take-up of the activities is estimated simultaneously. The evaluation of this probability involves three dimensional integration. Maximum likelihood estimations using numerical integration is only possible for a limited number of alternatives making the grouping of alternative diversification activities necessary. Formally the diversification choices can be presented as follows (Hardy and Bryman, 2004, p. 282):

$$y_{ij}^{*} = \begin{cases} y^{*}_{broadeing} & = \beta_{i1}x_{i1} + \varepsilon_{i1} \\ y^{*}_{deepening} & = \beta_{i2}x_{i2} + \varepsilon_{i2} \\ y^{*}_{combination\ of\ activities} & = \beta_{i3}x_{i3} + \varepsilon_{i3} \end{cases} \quad \text{and} \quad y_i = \begin{cases} 1 & if\ y_{ij}^{*} \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The probability of each farmer to choose the j-th alternative can be written as $P(y_i = j) = P_{ij} = P[\varepsilon_{i1} - \varepsilon_{ij} < (x_i - x_{i1})'\beta \ldots \varepsilon_{i4} - \varepsilon_{ij} < (x_i - x_{i4})'\beta]$.

In order to estimate only the activities chosen by the farmer the model restricts the diagonal axis of the variance- covariance matrix to ones. This avoids an identification problem. The joint distribution of all error terms is assumed to be multivariate normal distributed with covariance matrix $\sum$:

$$\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{bmatrix} \sim N(0, \sum)\ with \sum = \begin{bmatrix} 1 & \cdot & \cdot \\ \sigma_{21} & 1 & \cdot \\ \sigma_{31} & \sigma_{32} & 1 \end{bmatrix}.$$

The specific variance–covariance matrix is shown in table C.8. Table C.9 shows the estimated coefficients of the constructed MNP model.

Table C. 8 Covariance matrix for multinomial probit model

	age	fem	econsize	incdep	past	arab	hort
age	1.0000						
fem	-0.0689	1.0000					
econsize	-0.1744	0.3236	1.0000				
incdep	-0.0804	0.0351	-0.0535	1.0000			
past	0.0632	-0.0874	-0.2026	0.0383	1.0000		
arab	0.0830	-0.0797	-0.1246	0.0407	-0.4737	1.0000	
hort	-0.1019	0.2108	0.2762	-0.0513	-0.4718	-0.2033	1.0000
mix	-0.0023	-0.0041	-0.0020	0.0353	-0.2441	-0.1052	-0.1047
popdens	0.0077	0.0314	0.0344	0.0034	-0.0245	-0.0410	0.1017
mc	-0.0390	0.0489	0.0869	0.0510	-0.1557	0.2097	0.0965
ps	0.0165	-0.0420	-0.0402	-0.0968	0.0923	-0.0887	-0.1894
fc	0.0133	-0.0013	-0.0341	0.0056	0.0308	-0.0698	0.0532
du	-0.0028	0.0091	0.0361	0.0075	-0.0777	-0.0450	0.1954
lh	0.0271	-0.0123	-0.0339	0.0707	-0.0321	0.0561	-0.0057

	mix	popdens	mc	ps	fc	du	lh
mix	1.0000						
popdens	-0.0159	1.0000					
mc	0.0109	0.0294	1.0000				
ps	0.0147	-0.1005	-0.6069	1.0000			
fc	-0.0014	0.0403	-0.2122	-0.3996	1.0000		
du	-0.0198	0.0625	-0.0854	-0.1608	-0.0562	1.0000	
lh	0.0261	0.0383	-0.0710	-0.1336	-0.0467	-0.0188	1.0000

Table C. 9 Multinomial probit estimations specific activity categories

Number of obs	= 68724	
Log likelihood	= -27319.273	Prob > chi2 = 0.000
Wald χ^2 (42)	= 23696.08	

Broadening	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
AGE	-0.009	0.002	-4.940	0.000	-0.012	-0.005
FEM	0.035	0.005	6.550	0.000	0.025	0.046
ECONSIZE	-0.021	0.006	-3.900	0.000	-0.032	-0.011
INCDEP	1.432	0.016	89.140	0.000	1.400	1.463
PAST	0.098	0.088	1.110	0.267	-0.075	0.271
ARAB	-0.751	0.101	-7.400	0.000	-0.950	-0.552
HORT	0.096	0.095	1.010	0.311	-0.090	0.282
MIX	0.580	0.103	5.610	0.000	0.377	0.782
POPDENS	0.000	0.002	-0.200	0.843	-0.004	0.003
MC	-0.318	0.076	-4.180	0.000	-0.467	-0.169
PS	-0.556	0.072	-7.770	0.000	-0.696	-0.416
FC	-0.312	0.082	-3.780	0.000	-0.473	-0.150
DU	0.076	0.126	0.600	0.549	-0.172	0.323
LH	0.462	0.121	3.830	0.000	0.226	0.699
_cons	-2.995	0.146	-20.470	0.000	-3.282	-2.708
Deepening	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
AGE	-0.004	0.001	-4.250	0.000	-0.006	-0.002
FEM	0.016	0.006	2.410	0.016	0.003	0.028
ECONSIZE	0.001	0.003	0.380	0.705	-0.004	0.006
INCDEP	1.428	0.011	129.810	0.000	1.407	1.450
PAST	0.710	0.048	14.910	0.000	0.617	0.804
ARAB	-0.332	0.055	-5.990	0.000	-0.440	-0.223
HORT	-0.977	0.062	-15.670	0.000	-1.099	-0.855
MIX	0.240	0.064	3.750	0.000	0.115	0.366
POPDENS	-0.007	0.002	-4.430	0.000	-0.010	-0.004
MC	-0.483	0.042	-11.560	0.000	-0.564	-0.401
PS	-0.896	0.039	-23.100	0.000	-0.973	-0.820
FC	-0.573	0.046	-12.540	0.000	-0.663	-0.484
DU	-0.177	0.088	-2.010	0.045	-0.351	-0.004
LH	0.141	0.080	1.760	0.079	-0.016	0.299
_cons	-1.857	0.079	-23.510	0.000	-2.012	-1.702
Broadening & Deepening	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
AGE	-0.006	0.001	-4.330	0.000	-0.009	-0.003
FEM	0.022	0.005	4.430	0.000	0.012	0.031
ECONSIZE	-0.016	0.003	-5.260	0.000	-0.021	-0.010
INCDEP	1.334	0.013	100.320	0.000	1.308	1.360
PAST	-0.582	0.060	-9.770	0.000	-0.698	-0.465
ARAB	-0.459	0.065	-7.050	0.000	-0.587	-0.331
HORT	0.408	0.061	6.750	0.000	0.290	0.527
MIX	0.356	0.072	4.950	0.000	0.215	0.497
POPDENS	0.000	0.001	-0.070	0.941	-0.003	0.003
MC	0.202	0.076	2.660	0.008	0.053	0.351
PS	-0.011	0.074	-0.150	0.884	-0.156	0.134
FC	0.125	0.081	1.540	0.123	-0.034	0.284
DU	-0.069	0.120	-0.580	0.565	-0.304	0.166
LH	0.475	0.119	4.000	0.000	0.243	0.708
_cons	-2.786	0.117	-23.900	0.000	-3.014	-2.557

(No diversification is the base outcome)

Goodness of fit

Goodness of fit measures used for the MNP model are similar to the goodness of fit measures introduced for the binomial logit model. The values for the measures of fit provided by the data analysing software used is displayed in table C.10. The Log-likelihood value of the full model is -27,319.269. The Likelihood ratio Wald χ^2 test reveals with 42 degrees of freedom a significant value of 23,696.102. The hypothesis that the probability for each category is the same can be rejected. The Count R^2 is not to be compared with the McFadden R^2 discussed earlier. This measure of fit transforms the continuous predicted variables into binomial variables, and then assesses the predictions as correct or incorrect. In the multinomial case it holds also for this measure of fit that its explanatory power is rather low. Consequently Daganzo (1979, p. 112) recommends that one has got to rely on the researchers good judgment rather than on the statistical fit.

Table C. 10 Measures of fit for multinomial probit

Log-Lik Full Model:	-27319.269	D(68679):	54638.538
Wald X2(42):	23696.102	Prob > X2:	0.000
Count R2:	0.844	Adj Count R2:	0.227
AIC:	0.796	AIC*n:	54728.538
BIC:	-710298.121	BIC':	-23228.288
BIC used by Stata:	55139.741	AIC used by Stata:	54728.538