Conditional income disparity between farm and non-farm households in the European Union: A longitudinal analysis

Maria Marino, Benedetto Rocchi and Simone Severini

Maria Marino (contact author: maria.marino@uab.cat), Department of Applied Economics, UAB; Barcelona, Spain

Benedetto Rocchi, Department of Economics and Management, University of Florence, Italy.

Simone Severini, Department of Agricultural and Forestry Sciences, Universita degli Studi della Tuscia, Viterbo, Italy.

Abstract

Government interventions in the agricultural sector have been historically justified by the existence of an income disparity between farmers and non-farmers. However, recent studies have found that such disparity is disappearing over time, particularly in the United States. This work offers the first longitudinal systematic assessment on the average income disparity between farm and non-farm units in the European Union, differentiating for Old and New Member States. Using the EU-SILC dataset, both broad (having some farm income) and narrow (living mainly on agriculture) farm households are compared with a general sample of non-farm households and a more restricted sample of self-employed non-farm households. To control for household observable characteristics and time-constant unobserved factors, we use a fixed effects regression. Results suggest that the farm/non-farm income disparity has disappeared in the European Union unless we compare narrow farm households with all non-farm households: in this case, the former are more likely to be better-off than the latter. A limited income disparity is found only in the case of New Member States for broad farm households only. Results are used to draw policy considerations regarding the role of CAP in supporting farm income.

Keywords: Agricultural Households, Income, EU-SILC, European Union, Common Agricultural Policy.
JEL Classification codes: D31, I31, Q12, Q18

This is an UNCORRECTED DRAFT of the paper :

Marino M., Rocchi B., Severini S.. Conditional income disparity between farm and non-farm households in the European Union: A longitudinal analysis. Journal of Agricultural Economics (JAGE). Accepted for publication on Dec. 18th, 2020.

1. Introduction

In 2018, the European Commission proposed a set of amendments to the Common Agricultural Policy (CAP) and confirmed its political will to "*ensure a fair standard of living to farmers"* by supporting the agriculture sector. While the CAP now pursues a large set of objectives which are also related to the sustainable management of natural resources and a balanced territorial development, the objective to "*support viable farm income and resilience across the EU territory…*" is maintained (European Commission 2018).

Government interventions in such a direction have been historically justified by the so-called "farm income problem" which claims, among other things, that the level of farm income is systematically lower than non-farm income (Gardner 1992). Nevertheless, many empirical studies have found evidence that the income disparity between farm and non-farm groups is disappearing over time (Mishra et al. 2002; Katchova 2008). While some researchers claim that this is due to the effectiveness of public interventions (Mishra and El-Osta 2008; Mishra et al. 2009), others instead point to more structural economic and social changes, including the increasing role of off-farm incomes (Findeis and Reddy 1987; Mishra and Sandretto, 2002). The latter phenomenon, in particular, motivates Hill's recommendation (Hill 2012) to analyse the farm/non-farm income disparity between households rather than firms. He also recommends to investigate the farm income problem by looking at different definitions of farm households and of the non-farm comparison group. Indeed, both Eurostat (2001) and the UN/FAO handbook (United Nations 2012) make the point that the above definitions are critical. Generally, two definitions of farm households (FH) are employed in this literature: the *broad* definition includes all households having some farm income, no matter how small; the *narrow* definition refers only to households whose farm income accounts for at least half of the whole household income. The comparison group, the non-farm households (NFH), may also vary and range from all households in the population to households living on self-employed incomes only. The latter comparison is motivated by the fact that farming is indeed a self-employed activity.

The income disparity between FH and NFH has been studied mostly in the USA (Mishra et al. 2002; El-Osta et al. 2007; Katchova 2008). Only few studies investigate the farm/non-farm income disparity in the European Union (EU) (de Frahan et al. 2017; Rocchi et al. 2012; Stefani et al. 2012; Rocchi et al. 2020), despite a large share (almost 40% in 2018) of the EU budget being used for economic support to EU farmers through the CAP. For this reason, the present work offers a systematic EU-wide assessment of the income differential between the income of farm and non-farm households and answers the following three main research questions: Does a significant

disparity in the income level between farm and non-farm households still exist in the EU after controlling for observable and unobservable characteristics? Does the level of income disparity depend on how farm households and control groups are defined? Do Old and New Member States differ in terms of extent of such income disparity?

The analysis adds to the current literature by tackling the following methodological challenges: i) employ the household as unit of analysis; ii) create several comparison groups to represent farm and non-farm households; iii) control for observed and unobserved heterogeneity, through a fixed effects estimation with complex survey weights. We do so by exploiting the EU Statistics on Income and Living Conditions (EU-SILC), a longitudinal micro-level dataset representative of the European population.

The paper is organised as follows: Section 2 contains a brief review of the literature. Section 3 proposes an empirical strategy to deal with some of the main challenges explained above. Section 4 presents the data and how the definition of farm and non-farm households are built. Section 5 comments the results of our estimations. Some conclusions, policy implications and suggestions for further studies conclude the paper in Section 6.

2. Background on farm – non-farm income disparity

In a seminal paper, Gardner (1992) sheds light on the "farm income problem" both from a theoretical and an empirical perspective. The problem refers to the fact that when one compares a farm with a non-farm unit, the former is likely to be worse off than the latter. Farmers' relative disadvantage is generally analysed from a multidimensional perspective. Hill (2012), for instance, identifies three dimensions: the disparity issue, which refers to the comparison of farm income with the earnings in other sectors; the poverty issue in relation to specific regions or farm sizes; the instability issue which refers to the variations of income over time. Among these dimensions, the farm/non-farm income disparity is still attracting the attentions of many scholars because of its policy relevance, and is the focus of the present work.

Although agricultural sector is supposed to be disadvantaged with respect to other sectors, Gardner (1992) shows a general convergence between farm and non-farm income levels around the late 1980s in the USA. In some cases, farm units appear to be even richer than non-farm units. These findings have been confirmed by other studies (Mishra et al. 2002; El-Osta et al. 2007; Katchova 2008). A similar trend of convergence has been detected also in the EU (e.g. Eurostat 2001; de Frahan et al. 2017), although the phenomenon has been analysed much less than in the USA. Some works have also investigated the farm income problem in single EU countries, finding

mixed evidence (Phimister et al. 2004; Stefani et al. 2012; Rocchi et al. 2012).

Analysis which focuses on the income disparity dimension has been affected by three methodological problems. First, there has been a certain disagreement on which farm units are important to consider (firm or households). Nevertheless, a recent consensus is emerging that the most suitable way to approach the farm/non-farm income disparity is by considering households, since they have a greater command over the consumption of goods and services and they can adopt diversification strategies to cope with the increasing instability and risks in the sector (Mishra et al. 2002; Hill 2012; Hill and Bradley 2015). Within this framework, the difference between the *broad* and the *narrow* definition of FH acquires its relevance since it allows for the distinction between all households engaged in farming, including those who do not have farming as a primary economic activity, and those households which are "living on farming", where farm income is their main source of income.

Second, farmers' wellbeing has traditionally been assessed using sector-specific surveys (e.g. the Farm Accountancy Data Network). However, to assess the income disparity, it is important to specify a non-farm comparison group which cannot be directly created using sector-specific surveys. To overcome this problem, and despite their well-known potential limitations (e.g. the questionable representativeness of the farm sample; the inaccuracy of reported income data; missing information on farm benefits or land ownership; differences in the countries' accounting system), general surveys have become in the last years the preferred tool to investigate the farm/non-farm income disparity. Scholars can indeed compare FH (broad or narrow) with the entire population of NFH and with NFH whose main source of income comes from self-employed work (Hil 1982; Hill, 1999, 2012; Mishra et al. 2002; Hopkins and Morehart 2004; Hill and Bradley 2015). The rationale of the latter approach is that farmers are indeed self-employed workers.

These comparisons shed light on relevant and different policy issues (Mishra et al. 2002). On the one hand, we can analyse social gaps between all households somehow involved in farming (broad FH) – often considered a potential disadvantaged group – and the rest of the population. On the other hand, we can analyse productivity gaps among sectors (agriculture vs other industries), by comparing households that are mainly living on agriculture (narrow FH) with households living on self-employed work in other industries.

Third, methodologies to test differences between farm and non-farm units need to take into account observable and unobservable heterogeneity. While scholars have generally employed parametric and nonparametric tests with regression analysis to account for differences in terms of human capital, age or other observable characteristics (Pender and Gebremedhin 2006; El-Osta et al. 2007; Mishra and Moss 2008; Katchova 2008), unobserved factors may still affect both selection

into the farm sector and income levels. For example, unobservable preferences might determine the sorting of households into the farm sector. Similarly, unobserved characteristics, such as skills, might affect incomes (e.g. Gasson 1973; Gorton et al. 2008; Berkhout et al. 2010; Howley and Dillon 2012). Despite the inter-sectorial literature, which generally focuses on wage differentials (e.g., between the public and private sector, formal and informal, etc.), has proposed different empirical strategies to address the issue (e.g. Bargain and Kwenda 2011; Nguyen et al. 2007; Lausev 2014; Christopoulou and Monastiriotis 2014; Hospido and Moral-Benito 2016), in the literature on the farm parity issue, unobserved heterogeneity has been rarely taken into account with very few exceptions (Stefani et al. 2012, De Frahan et al. 2017; Rocchi et al. 2020). Thanks to the rotating panel nature of the EU-SILC, the present work exploits the longitudinal structure and employs fixed effects regressions, which allow individual unobserved characteristics to be net out. Although fixed-effects regressions are not a panacea against all biases caused by omitted variables, they are very useful in removing the potential and structural determinants of both being a farmer and their income.

3. Estimation strategy

We consider the household as unit of analysis. More specifically, denote with y_{it} the level of equivalised disposable income of households i at time t. We can estimate a standard income equation (Mincer, 1958):

$$y_{it} = \alpha_i + \gamma_t + x_{it}\boldsymbol{\beta} + FH_{it}\boldsymbol{\delta} + \varepsilon_{it} \qquad (1)$$

where α_i is the household fixed effect; γ_t collects year fixed effects; x_{it} are the control variables, FH_{it} represents a dichotomous variable taking the value of 1 if the i-th observation at time t is a farm household, and ε_{it} is a classical Gaussian error.

Unlike its natural competitor, the random effects model whose unobserved α_i is assumed to be independent of x_{it}, the fixed effects (FE) model allows α_i to be correlated with the control variables. This allows for the consistent estimation of partial effects in the presence of time-constant omitted variables which can arbitrarily be related to the observables x_{it}. Despite FE models are generally standard in the literature of inter-sectorial income gaps, we also run a robust Hausman test in Section 5 (Wooldridge 2010) to give strength to our choice of fixed over random effects.

Given the complex survey design of the EU-SILC, we can partially address the violations of the homoscedasticity and independence by including sample weights in our estimation. The

literature on complex surveys has indeed shown that the omission of sample weights produces a serious underestimation of standard errors (Kish 1995; Lohr 2000). Moreover, Kott (1991) shows that weighted estimates are more robust to omitted variable problems and to heteroscedasticity. To control instead for serial correlation, we estimate coefficients whose standard errors are clustered at household level, which is the unit more likely to be correlated over time (Arellano 1987; Woolridge 2010).

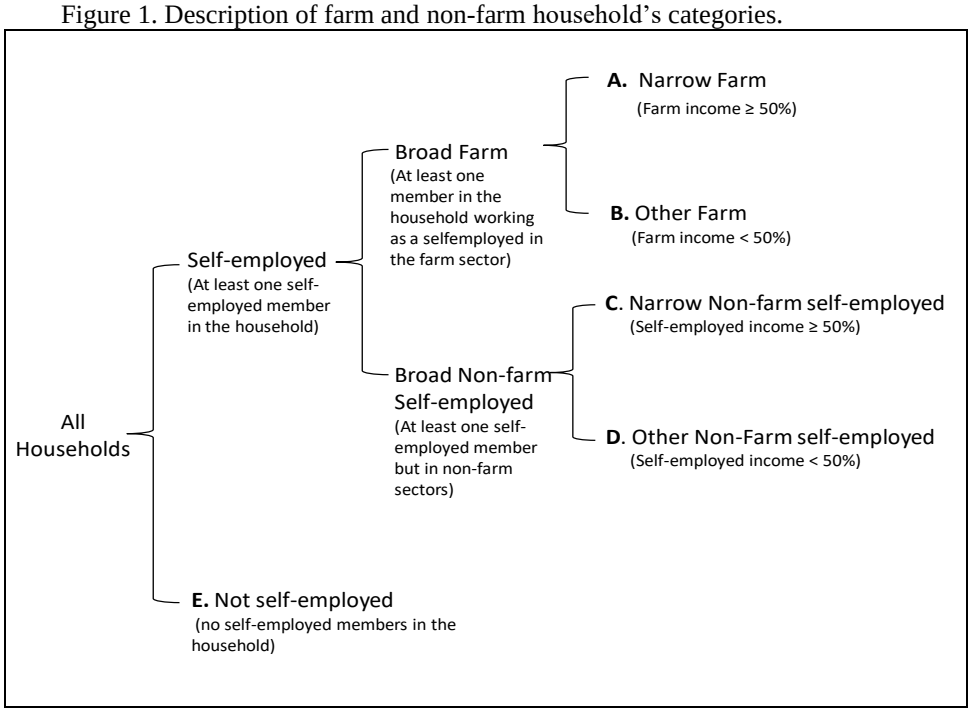
Moreover, to overcome the problem of outliers and influential observations, we recalibrate sample weights following the approach proposed by Alfons et al. (2013) and generally adopted by those working with income variables (Alfons and Templ 2013; Brzesinki 2016; Jenkins 2017; Safari et al., 2018; Templ et al., 2019; Safari et al. 2019). This procedure consists of detecting outlier observations against a fitted Pareto distribution of the variable of interest, applying the Van Kerm's rule of thumb to determine the threshold (Kerm 2007). The weights of non-representative outliers are set to 1 while the weights of the remaining observations are adjusted accordingly by calibration, leaving the total sum of weights unchanged. We perform the recalibration within each country/year group of observations.

Finally, the problems of zero (or negative) values as well as the skewness of the income variable are addressed by using a transformation through the inverse hyperbolic sine (IHS) (Burbidge et al. 1988).

To conclude on the estimation strategy, we run several specifications. First, we compare a restricted base model (RM) of Equation (1) with an unrestricted model (UM). The difference between the two is that the latter adds to the former the interaction terms between the x_{it} and the farm household dichotomous variable. Such specification aims at addressing potential issues arising from the assumption that returns to the control variables are identical across farm and non-farm households. In Section 5, we test such assumption by comparing the model with and without interactions through a Wald test. The latter checks whether the parameters of interest (the interactions) are simultaneously equal to zero. If they are, this strongly suggests that removing them from the model will not substantially reduce the fit of the model (Chambers and Skinner 2003).

Second, the focus of our analysis is the estimated coefficient δ on FH_{it} in Equation (1) since it quantifies the conditional income disparity between farm households (FH) and non-farm households (NFH), with the latter group representing our base category. However, we estimate different coefficients δs according to alternative definitions of farm households (FH) and non-farm households (NFH). In our analysis, we consider three FH-NFH comparisons which give rise to different specifications. To better understand, Figure 1 shows how the sample of the entire population (All Households) is split.

Figure 1. Description of farm and non-farm household's categories.



We first identify self-employed households, namely those who have at least one self-employed member. The rest of the sample consists of non-self-employed households, namely families with no self-employed members (letter E). Within the self-employed households, we identify broad farm households when at least one member is self-employed and works in the agricultural sector. The rest of the self-employed households, the broad non-farm self-employed group, have at least one self-employed member but none working in the farm sector. The broad farm households are then split between narrow farm households, whose main source of income (at least 50% of the total household income) is coming from self-employed work in agriculture (letter A) and, residually, the rest of farm households (letter B). Finally, broad self-employed non-farm households are also split between the narrow self-employed non-farm households, whose main source of income (at least 50%) is coming from self-employed work (letter C) and, residually, the rest of the self-employed non-farm households (letter D).

Table 1 shows our three specifications. Specifications C1 and C2 compare the income of broad and narrow farm households with that of all non-farm households (NFH). Specification C3 compares narrow farm households (A) with narrow self-employed non-farm households (NFH*) whose main source of income comes from self-employed work (C).

Table 1. Specification of the income disparity problem and comparison groups

Specifications	Comparison groups (FH vs. NFH)		
C1	A + B	vs.	C + D + E
C2	A	vs.	C + D + E
C3	A	vs.	C

According to the three specifications, the base category changes as well as the samples, so the estimated coefficient δ of FH_{it} assumes different meanings. Specifications C1 and C2 tell us how much (broad or narrow) farmers are better or worse-off with respect to the living standards of the (non-farm) rest of society. In contrast, coefficient δ of C3 provides an estimate of the income disparity between narrow farm households and narrow self-employed non-farm households. Interestingly, the last specification allows for interpretation of δ as a closer approximation of a sector effect, provided that the comparison is between two groups where (farm or non-farm) self-employed income is more than a half of the overall household income.

Finally, for the three above specifications of the parity problem, in addition to the standard measure of household equivalised disposable income, we re-run the model using an extended version of the household income including also non-monetary sources of income.

4. Data

Our data source is EU-SILC, a survey based on a mix of register sources and direct interviews which offers a wide range of monetary and non-monetary variables. The economic activity and the income received in the reference year are collected for each member of the family who is more than 15 years old. The availability of detailed information on incomes of the household allows the direct application of the standard concepts proposed by the Wye Group Handbook (United Nations 2012: 449-452) for the identification of agricultural households.

We employ the longitudinal version of the data which ranges from 2004 to 2015 with a rotating structure that covers the same households over 4 years.[1]

Even though EU-SILC includes all European countries, our analysis restricts the sample only to countries belonging to the European Union. Since disregarding differences among Member States can be misleading, in our estimates we also distinguish between two groups of countries according to the time of entry in the European Union: New and Old member states (NMS and

1 Rotating structure refers to the sample selection based on a number of subsamples or replications. Some replications are retained, while others are dropped and replaced by new replications. In particular, EU-SILC chooses four replications. Any particular replication remains in the survey for four years; each year, one of the four replications from the previous year is dropped and a new one added. More details on the sample design of the EU-SILC survey can be found on the Eurostat institutional website at the URL https://ec.europa.eu/eurostat/web/income-and-living-conditions/data.

OMS). Only 22 out of 28 countries of the European Union are included in our empirical analysis since variables which enable us to define farm households are missing. Our sample includes: Austria, Belgium, Denmark, Finland, France, Greece, Italy, Portugal, Spain, Sweden and the United Kingdom (OMS), as well as Bulgaria, Cyprus, Check Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania and Slovakia (NMS).

The EU-SILC database does not directly provide a farm/non-farm classification as provided by Figure 1. Consequently, we first distinguish self-employed individuals from those who are not, since farm income is generated by self-employment work. Among the self-employed individuals, we assign a dichotomous variable taking the value of 1 when the individual works in the agricultural sector according to the ISCO-88 classification. Once farmers are identified, broad farm households (A + B) are defined by selecting households with at least one farmer. Residually, the broad self-employed non-farm households (C + D) are those in which there is at least one non-farmer self-employed individual.

To consider the narrow definition, we need first to compute farm incomes at the family level by summing up all self-employment incomes of members classified as farmers. Then, within the broad group, narrow FH (A) are identified by selecting families earning a total farm income that is at least half of the total disposable household income. In the same way, narrow self-employed non-farm households (C) are identified by selecting non-farm families whose self-employment income is more than the half of the total disposable income. The EU-SILC variables selected to perform the empirical analysis are reported in Table A1.

Table 2 reports the number of FH and NFH observations along the whole period and for the total sample (EU), as well as for the two subsamples referring to the NMS and OMS. On average, the yearly total number of EU families represented by the total sample is about 202 million. Represented FH are respectively about 6 million according to the broad definition and 2 million according to the narrow definition. Table A2 in the Appendix contains the number of the observations for all countries included in our panel dataset. Note that, despite some countries having a limited number of observations, the total number in the EU, NMS and OMS samples is suitable to support the empirical exercise for the whole period considered.

Table 2. Sample size (2004 – 2015) by groups of countries and household comparison groups.

European Union	Farm Households (FH Broad: A+B; Narrow: A)	44 663	15 685
	Non-farm Households (NFH: C + D + E)	1 404 664	
	- of which:		
	Narrow Non-farm Self-employed (NFH: C)*	*70 119*	
New Member States	Farm Households (FH) (Broad: A+B; Narrow: A)	26 092	8 302
	Non-farm Households (NFH: C + D + E)	593 023	
	- of which:		
	Narrow Non-farm Self-employed (NFH: C)*	*49 274*	
Old Member States	Farm Households (FH) (Broad: A+B; Narrow: A)	18 571	7 383
	Non-farm Households (NFH: C + D + E)	811 641	
	- of which:		
	Narrow Non-farm Self-employed (NFH: C)*	*49 274*	

Source: Own elaborations on EU SILC data.

The dependent variables are the level of equivalised disposable income and an extended definition of income which includes also non-monetary sources, namely the imputed rents from property or land and the in-kind incomes from self-consumption of produced goods. All income variables are adjusted by using purchasing power parities (PPPs). More specifically, we consider the real values of the income variable from the price level indices and real expenditures for the European Standard of Accounts (2010 aggregates - Eurostat).[2]

Finally, to take into account observable heterogeneity, the model includes a standard set of controls generally employed among the determinants of household income (De Janvry and Sadoulet 2001; Mendola 2007; Becerril and Abdulai 2010). Some of them define the characteristics of the householder (age, age squared, education, marital status, gender, health)[3], while others define household's features (number of components and region of living).[4]

All independent and dependent variables are described in Table 3.

[2] We take these data from Eurostat (http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=prc_ppp_ind&lang=en . Accessed on 14/05/2019.

[3] The householder is defined using the following criteria: first, he/she must be responsible of the accommodation; second, if there are two householders, the one earning the greatest income is considered; finally, if still no householder is identified, we consider the eldest.

[4] We test for the presence of multicollinearity using the variance inflation factor, but we do not detect such a problem in our data.

Table 3. Descriptive statistics and definition of dependent and independent variables.
EU, NMS and OMS means estimated by complex survey weights (Standard Deviation in parentheses)

Variable (label)	EU	NMS	OMS	Description
INCOME	17,1	8,396	19,624	Real Equivalised Disposable household income (PPP)
	(13,474.47)	(6,853.93)	(13,865.17)	
EXTENDED INCOME	19,121	9,637	21,871	Real Equivalised Disposable household income and in-kind and imputed rent (PPP)
	(14,546.68)	(7,365.4)	(14,952.94)	
Age (agehead)	51	50.61	51.11	Age of the household holder (35-55)
	(16.86)	(16.54)	(16.95)	
Middle educated (mided)	0.55	0.64	0.52	Education of the household holder 1=lower secondary education & upper secondary education
	(0.5)	(0.48)	(0.5)	
Highly educated (highed)	0.26	0.24	0.27	Education of the household holder 1=post-secondary non tertiary education & tertiary education
	(0.44)	(0.43)	(0.44)	
Married (mrdhead)	0.35	0.39	0.34	Civil status of the household holder 0=not married; 1= married
	(0.48)	(0.49)	(0.47)	
Separated (sprhead)	0.1	0.09	0.1	Civil status of the household holder 0=not married; 1= separated OR divorced
	(0.3)	(0.29)	(0.31)	
Widowed (widhead)	0.13	0.17	0.12	Civil status of the household holder 0=not married; 1= widowed
	(0.34)	(0.37)	(0.33)	
Sex (malehead)	0.61	0.59	0.61	Sex of the household holder 0=female; 1= male
	(0.49)	(0.49)	(0.49)	
Health (healhead)	1.54	1.40	1.58	Self-reported health of the household holder 0=bad and very bad; 1=fair; 2= good and very good
	(0.68)	(0.74)	(0.66)	
Residence (rural)	0.26	0.45	0.21	Living in a rural area 0= urban or suburbs; 1 = rural
	(0.44)	(0.5)	(0.4)	
Size (hh_size)	2.45	2.69	2.39	Number of individuals living in the household
	(1.35)	(1.52)	(1.29)	

Source: Own elaborations on EU SILC data.

5. Results

5.1. Income levels and characteristics of the households

According to the raw (i.e. unadjusted) average income levels, farm households, whichever definition is used, are always worse off than the relative comparison groups (Table 4). Although not reported for space reasons, all income differences between farm and non-farm groups are statistically different from zero.[5]

In the EU, FH incomes are on average around half of those earned by NFH. When looking to the two subsamples, the picture does not change substantially: FH incomes are much lower in the NMS than in OMS, for both income definitions. Broad FH are generally poorer than narrow FH, but the differences are not striking. Although beyond the scope of the paper, coefficients of variation suggest that incomes distributions are quite skewed in the EU, especially when looking at narrow FH. In the NMS and the OMS, however, narrow self-employed NFH have a higher coefficient of variation.[6]

Table 4 Equivalised income levels in the EU, NMS and OMS sample (Euro)
Average mean estimated by complex survey weights (Standard Deviation in parentheses).

	Broad FH (A + B)	Narrow FH (B)	NFH (C + D + E)	NFH* (C)
		European Union		
Income	9,465.02	10,552.01	17,304.52	22,056.98
	(10,641.37)	(13,051.02)	(13,636.7)	(24,241.08)
Extended income	10,939.65	12,210.6	19,331.19	24,348.6
	(11,244.54)	(13,713.08)	(14,241.72)	(24,348.6)
		New Member States		
Income	4,818.30	5,107.68	8,637.85	11,538.77
	(4,622.84)	(6,405.99)	(9,561.7)	(15,410.46)
Extended income	5,887.12	6,298.73	9,890.85	12,900.34
	(5,106.76)	(7,064.08)	(10,284.57)	(15,874.22)
		Old Member States		
Income	15,114.74	15,859.32	19,690.14	24,332.11
	(12,129.92)	(14,481.85)	(12,079.61)	(23,684.15)
Extended income	17,082.77	17,973.69	21,929.78	26,824.89
	(12,566.17)	(14,374.29)	(12,516.8)	(14,917.45)

Source: Own elaborations on EU SILC data.

The comparison of raw income means can be misleading because part of the disparity might be due to households' differences in observable and unobservable characteristics. Indeed, FH and NFH groups deeply differ when considering observable characteristics (Table 5). In particular, FH are less educated (*highed*), are more likely to live in rural areas (*rural*), have larger family size (*hh_size*) and it is more likely that the household head is male, among other things.

5 Incomes level are estimated with the STATA command *svy:mean* and the differences are tested using the command *lincom*. Test differences are available upon request.

6 Such insights on income disparity within household categories are supported also by the estimation of Gini indexes and the ratios between the average income of the highest and lowest decile. These additional figures on income inequality are available upon request to the authors.

Table 5 also reports the significance test of the differences among the FH and NFH independent variables in the whole EU sample. In the Appendix we also report the differences and their relative test for the NMS and OMS (Table A3) and the picture is very similar to the whole sample. We also test whether income differences exist between groups stratified according to observable characteristics and type of household (farm and non-farm), and we find that there are statistically significant income disparities between these groups in almost all cases.[7]

Table 5 Mean and differences of the independent variables by household group estimated by using complex survey weights. Whole EU sample (Standard Deviation in parentheses)

Variables		C1 Broad FH vs. NFH		C2 Narrow FH vs. NFH		C3 Narrow FH vs. NFH*	
agehead	NFH	51.15	(16.89)	51.15	(16.91)	44.74	(10.3)
	FH	45.57	(14.17)	45.8	(11.45)	45.8	(11.86)
	Diff.	5.58	***	5.35	***	-1.07	***
mided	NFH	0.55	(0.5)	0.55	(0.5)	0.6	(0.48)
	FH	0.68	(0.5)	0.72	(0.49)	0.72	(0.51)
	Diff.	-0.13	***	-0.17	***	-0.12	***
highed	NFH	0.27	(0.44)	0.27	(0.44)	0.31	(0.45)
	FH	0.11	(0.33)	0.07	(0.28)	0.07	(0.29)
	Diff.	0.16	***	0.2	***	0.24	***
mrdhead	NFH	0.35	(0.48)	0.35	(0.48)	0.5	(0.49)
	FH	0.46	(0.53)	0.58	(0.54)	0.58	(0.56)
	Diff.	-0.11	***	-0.23	***	-0.08	***
sprhead	NFH	0.1	(0.3)	0.1	(0.3)	0.09	(0.28)
	FH	0.03	(0.19)	0.05	(0.23)	0.05	(0.24)
	Diff.	0.07	***	0.06	***	0.04	***
wdwhead	NFH	0.14	(0.34)	0.14	(0.34)	0.02	(0.14)
	FH	0.05	(0.22)	0.03	(0.18)	0.03	(0.18)
	Diff.	0.09	***	0.11	***	-0.01	***
malehead	NFH	0.61	(0.49)	0.61	(0.49)	0.79	(0.4)
	FH	0.74	(0.47)	0.87	(0.37)	0.87	(0.38)
	Diff.	-0.13	***	-0.27	***	-0.08	***
healhead2	NFH	1.53	(0.69)	1.53	(0.69)	1.77	(0.47)
	FH	1.66	(0.63)	1.72	(0.57)	1.72	(0.59)
	Diff.	-0.12	***	-0.18	***	0.06	***
rural	NFH	0.25	(0.43)	0.25	(0.43)	0.25	(0.42)
	FH	0.78	(0.45)	0.79	(0.45)	0.79	(0.46)
	Diff.	-0.53	***	-0.54	***	-0.54	***
hh_size	NFH	2.42	(1.33)	2.42	(1.33)	2.92	(1.3)
	FH	3.56	(1.79)	3.3	(1.72)	3.3	(1.78)
	Diff.	-1.14	***	-0.88	***	-0.38	***

Source: Own elaborations on EU SILC data.

[7] Results of the tests are not reported for space reasons, but they are available upon request. Generally, in the C1 comparison, all income differences are statistically significant except when the size of family increases. This is similar for C2 and C3.

5.2 Conditional farm/non-farm income differences

The presence of the above statistically significant differences suggests the importance of controlling for observables. However, as explained in Section 3, it is also crucial to take unobservable (time invariant) characteristics of the households into account in order to get consistent estimates of income differentials. In this Section, we consider the three specifications of the income disparity problem under which we estimate the coefficient δ (Table 1).

Before estimating the fixed effects models for the EU, NMS and OMS, the standard practice in the panel regression literature asks to choose between random and fixed effect model (Hausman, 1978). Given that the standard Hausman test requires the homoscedasticity assumption to be respected, we adopted an artificial regression approach as suggested by Arellano (1993) and Wooldridge (2010) which is a generalisation of the Hausman test. More specifically, we use a Sargan-Hansen statistic that is robust to heteroscedasticity and within-group correlation.[8] Table A4 in the Appendix reports the test and the p-value. As shown by the critical values, the null hypothesis of no correlation is always rejected supporting the use of FE.

A Wald test is then used to choose between a model's specification without or with interactions (RM vs UM) among the covariates and the farm household dichotomous variable. Table A5 reports the Wald test[9] and the p-values showing that the null hypothesis that the interactions are jointly equal to zero is not rejected in most cases (or in all cases using 1 per cent as cut-off for statistical significance). This indicates that the restricted model is to be preferred since including these variables does not create a statistically significant improvement in the fit of the model. In other words, covariates' coefficients do not differ across the two groups (Table A5).

Tables A6-A8 in the Appendix report the estimation results for C1, C2 and C3 for the whole EU, NMS and OMS. Most coefficients on the independent variables have the expected sign and confirm the importance of variables such as the level of education and the health of the householder. Although not reported, years dummies show significant and positive trends in almost all models' specifications. Table 6 reports the δ coefficients (the conditional farm/non-farm income differences) for the three samples (EU, NMS and OMS) and the three specifications of model (1) (C1, C2, C3).

A negative income disparity between farm and non-farm households exists when considering broad FH in both the whole EU and NMS-OMS sub-samples. However, the coefficients

[8] The robust Hausman test is based on Schaffer and Stillman (2006) STATA user written command in *xtoverid*, which is recommended when robust standard errors are used. The command is executed after the estimation of a random effect models (STATA command, *xtreg, re vce(cluster))*.

[9] Tests performed using the STATA command *test* on the interaction variables after the estimation of the unrestricted model containing all interactions.

are almost always not significant, except when considering the NMS, where FH are worse off than the rest of the population (C1). Such results confirm the hypothesis that, on average, farm households filled the income disparity with the rest of society but progresses still need to be achieved in those countries which are relatively new in the EU (Gardner, 1992; Hill and Bradley, 2015; De Frahan et al., 2017). Note that the finding of a convergence between broad FH and all other households holds also when considering non-monetary sources of income and this is in line with most of the literature cited in Section 1 and 2.

When exploring the income disparity between narrow FH and NFH (C2), the picture changes. In particular, narrow farm families are better-off on average than non-farm households in all samples except when considering the extended income in the NMS, where the income disparity is not significant. Such results are in line with Eurostat (2001) but partly contradict the last evidence in de Frahan et al. (2017), although the latter focuses on OECD countries only.

Finally, considering the comparison with the group of narrow self-employed non-farm households (C3), narrow FH incomes are not different from their counterpart in all samples. Such comparison has not been proposed so far in the literature despite providing important insights. First, it reduces the differences in terms of observable characteristics and possibly unobservable characteristics related to the choice of self-employed work. Moreover, it gives a much clearer picture on the difference between agriculture and other industries in the returns of self-employed income (see, for example, Mishra et al. 2002; Marshall and Peake 2009 for similar works). In particular, C3 reveals that a sectorial gap between agriculture and other sectors does not exist when the comparison group refers to only households living on self-employment.

Table 6. Income disparity between farm and non-farm households (δ coefficient in (1))
Broad and narrow definitions of Farm Households.
EU, NMS and OMS sample. (Clust. Rob. Standard Errors in parentheses)

	Farm Households vs. Non-farm households		Narrow Farm Households vs. Narrow Self-employed Non-farm households
	C1	C2	C3
		European Union	
Income	-0.0354	0.2820***	-0.1104
	(0.0321)	(0.0716)	-0.0716
Extended Income	-0.0206	0.1706***	-0.0867*
	(0.0205)	(0.0468)	(0.0493)
		New Member States	
Income	-0.0590**	0.1799**	-0.0723
	(0.0251)	(0.0897)	(0.0632)
Extended Income	-0.0487**	0.1202	-0.0530
	(0.0215)	(0.0819)	(0.0545)
		Old Member States	
Income	-0.0264	0.3294***	-0.1283
	(0.0522)	(0.0958)	(0.1093)
Extended Income	-0.0099	0.1944***	-0.1054
	(0.0315)	(0.0564)	(0.0708)

* p < 0.10, ** p < 0.05, *** p < 0.01.
Source: Own elaborations on EU SILC data.

5.3 Robustness test

We assess how robust are our results considering three aspects.

The first aims at excluding that our results are driven by incidental changes in the professional status like retirement. For such reason, we re-estimate Table 6 by focusing only on the sample of those households whose householder is less than 65-year-old. Results reported in Table B1 shows that this is not the case.

The second develops the same estimations but restricting our sample to the country/year groups of observations including at least 30 (broad) FH (de Frahan et al., 2017). In such a way, we end up with less countries and years, but such restrictions allow us to check if our previous results are driven by random variation in the data on farm households. More specifically, we end up with 14 countries. Table B2 reports the results which show that the income disparities estimated on the restricted sample are in-line with those described in Table 6. This is true for all three comparisons we developed (i.e. C1, C2 and C3) and the two variants of income (i.e. income and extended income). The only exception is for specification C2 in the sample of the NMS where the estimated coefficient is now not significant.

The third aspect refers to the fact that histories as well as social, economic and legal structures are so different among EU members that only comparisons between farm and non-farm household incomes within individual countries could seem relevant and safe to interpret. We are

strongly convinced that our EU-wide estimates are relevant for researchers and for policy makers who are interested in the general picture of the farm income problem. Nonetheless, we exploited the EU-SILC dataset to assess if relevant differences could exist in FH - NFH income disparities among single countries. Since EU-SILC hinders the possibility to do so for a large number of countries, we developed the estimations for Finland, France, Greece, Italy and Spain for the OMS, and Lithuania, Poland and Romania for the NMS. Table B3 shows that aggregate results of Table 6 are in line with single countries. In the OMS, the income differential estimates between the broad farm households and non-farm population (C1) are not significant as in the case of our base estimates reported in Table 6. Similarly, the income differential between narrow FH and the whole non-farm population (C2) is significant and positive as in the case of the whole OMS sample in three out of five countries. Finally, the income differentials between narrow FH and narrow self-employed non-farm households (C3) is coherent with that of the whole OMS sample in four out of five cases, with the noticeable exception of Finland where the income differential is significant and negative. Similar considerations apply to the three selected NMS. Coherently with Table B1, results for the three countries support the main picture: two (Lithuania and Poland) out of three countries do not show any income disparities while Romania records a negative and significant income disparity. The results for the income disparity between narrow FH and all NFH (C2) are also sufficiently in line with those reported in Table B1. Finally, the results about the income disparity between narrow FH and narrow self-employed NFH (C3) of the considered three countries are fully consistent with those obtained for the whole NMS sample. In the end, the results of individual country estimates suggest that, while some inter-country variations exist, these are not so large to confound the interpretation of the average income differential at the three considered levels of aggregation (EU, OMS, NMS).

6. Conclusions

The present analysis evaluates the income disparity between farm and non-farm households in the European Union using data from the European Union Survey on Incomes and Living Conditions (EU-SILC) in the period 2004-2015. The analysis refers to the whole EU as well as to two sub-groups of countries (Old and New Member States), and identifies farm and non-farm households according to different criteria, considering not only monetary income but also non-monetary sources of income. Furthermore, differing from previous studies referring to the EU, we condition the estimates of the income disparity to differences both in observables and unobservable characteristics of households that are likely to affect the income level.

The analysis shows that the level of income disparity depends on how farm households and control groups are defined. According to our results, on average, farmers can no longer be considered as a homogeneous social group and a relevant difference exists between *broad* and *narrow* farm households. *Broad* farm households show an income level comparable with the rest of the population. Furthermore, *narrow* farm households have an average income level *above* the total non-farm population conditional average. However, the income of *narrow* farm households is similar to that of other self-employed non-farm families. In other words, being self-employed in agriculture instead of in other industries no longer represents a relative disadvantage across the EU countries. A similar picture emerges when considering the OMS, while a limited income disparity is found in the case of NMS.

The results allow to draw some relevant policy implications. Most of the CAP support is channeled through decoupled direct payments (DDP) that are essentially directed to support the income of all farm households (including also families for which farming is only a secondary source of income) (Biagini et al., 2020). However, this goal is achieved at the cost of generating an income advantage to the *narrow* farm families, whose income level is higher than the average of non-farm population, especially in the OMS. These results, considering that the largest beneficiaries of decoupled direct payments are likely to be included in the *narrow* group of FH (Matthews, 2016; Severini and Tantari, 2015), suggest to question whether these payments are effectively targeted to support the farmers who experience a relevant income gap.

The analysis presented here has some limitations that are worth being highlighted. First, EU-SILC is a survey representative of the totality of households in the EU, but farm households represent only a limited group. Consequently, it may be under-represented (United Nations 2012: 588). Second, our analysis focuses on average values only. Future extensions, aimed at exploring the whole distribution of the sample or the evolution over the whole period, could provide additional insights. A further limitation refers to our estimation strategy: a fixed effect estimation is only a partial solution to endogeneity since time-varying factors and selection issues remain. Nevertheless, finding instruments in our context of analysis is complex given the nature of our dataset. Moreover, fixed-effect models explore only within households variation. While we have a sufficient number of changes in professional status from farm to non-farm household and vice versa and these are not systematically driven by retirement, our dataset does not allow for a more sophisticated analysis. Nonetheless, our analysis still represents an advance on the topic since, to the best of our knowledge, it is the very first time that income disparities between farm and non-farm households are estimated in a longitudinal setting.

References

Alfons, A., Templ, M. and Filzmoser P. (2013). Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 62 (2): 271-286.

Alfons, A., Templ, M. (2013). Estimation of social exclusion indicators from complex surveys: the R package *laeken*. *Journal of Statistical Software*, 54 (15): 1-25.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. Oxford bulletin of Economics and Statistics, 49(4), 431-434.

Arellano, M. (1993). On the testing of correlated effects with panel data. Journal of econometrics, 59(1-2), 87-97.

Bargain, O., and Kwenda P. (2011). Earnings Structures, Informal Employment, and Self-Employment: New Evidence From Brazil, Mexico, and South Africa. *Review of income and wealth*, 57:100-122.

Becerril, J., & Abdulai, A. (2010). The impact of improved maize varieties on poverty in Mexico: a propensity score-matching approach. *World development*, 38(7), 1024-1035.

Berkhout, E. D., Schipper, R. A., Kuyvenhoven, A., & Coulibaly, O. (2010). Does heterogeneity in goals and preferences affect efficiency? A case study of farm households in northern Nigeria. *Agricultural Economics*, 41(3/4), 265-273.

Biagini, L., Antonioli, F. & Severini, S. (2020). The Role of the Common Agricultural Policy in Enhancing Farm Income: A Dynamic Panel Analysis Accounting for Farm Size in Italy. *Journal of Agricultural Economics* (JAGE), 71 (3), September 2020: 652-675.

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83(401), 123-127.

Brzesinski, M. (2016). Robust estimator of the Pareto tail index: a Monte Carlo analysis. *Empirical Economics*, 51(1): 1-30.

Chambers, R. L., & Skinner, C. J. (Eds.). (2003). Analysis of survey data. John Wiley & Sons.

Christopoulou, R., and Monastiriotis V. (2014). The Greek Public Sector Wage Premium before the Crisis: Size, Selection and Relative Valuation of Characteristics. *British Journal of Industrial Relations* 52-53: 579-602.

de Frahan, B. H., Dong, J., & De Blander, R. (2017). Farm household incomes in OECD member countries over the last 30 years of public support. In *Public Policy in Agriculture* (pp. 148-175). Routledge.

De Janvry, A., & Sadoulet, E. (2001). Income strategies among rural households in Mexico: The role of off-farm activities. *World development*, 29(3), 467-480.

El Osta, H. S., Mishra, A. K., & Morehart, M. J. (2007). Determinants of economic well-being among US farm operator households. *Agricultural Economics*, 36(3), 291-304.

European Commission. (2018) 'Proposal for a regulation of the European Parliament and of the Council establishing rules on support for strategic plans to be drawn up by Member States under the Common agricultural policy (CAP Strategic Plans) and financed by the European Agricultural'. Available at https://eur-lex.europa.eu/resource.html?uri=cellar:aa85fa9a-65a0-11e8-ab9c-01aa75ed71a1.0003.02/DOC_1&format=PDF

Eurostat. (2001). Income of Agricultural Households Sector - 2001 Report (CD-ROM) - Product - Eurostat. https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-46-02-953-

3A?inheritRedirect=true. (last accessed 14 may 2019).

Findeis, J. L., & Reddy, V. K. (1987). Decomposition of income distribution among farm families. *Northeastern Journal of Agricultural and Resource Economics*, 16(2), 165-173.

Gardner, B. L. (1992). Changing economic perspectives on the farm problem. *Journal of economic literature*, 30(1), 62-101.

Gasson, R. (1973). Goals and values of farmers. *Journal of agricultural economics*, 24(3), 521-542.

Gorton, M., Douarin, E., Davidova, S., & Latruffe, L. (2008). Attitudes to agricultural policy and farming futures in the context of the 2003 CAP reform: A comparison of farmers in selected established and new Member States. *Journal of Rural Studies*, 24(3), 322-336.

Hausman, J. A. (1978). Specification tests in econometrics. Econometrica: Journal of the econometric society, 1251-1271.

Hill, B. (1982). Concepts and measurement of the incomes, wealth and economic well-being of farmers. Journal of Agricultural Economics, 33(3), 311-324.

Hill, B. (1999). Farm household incomes: perceptions and statistics. *Journal of rural studies*, 15(3), 345-358.

Hill, B. (2012). Farm incomes, wealth and agricultural policy filling the CAP's core information gap. Cabi.

Hill, B., and Bradley B. D. (2015). Comparison of Farmers' Incomes in the EU Member States http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2015)540374. (last accessed 14 may 2019).

Hopkins, J., & Morehart, M. (2004). Determinants of Poverty among Us farm households. *Amber Waves*, 2 (1): 8.

Hospido, L., & Moral-Benito, E. (2016). The public sector wage premium in Spain: evidence from longitudinal administrative data. *Labour Economics*, 42, 101-122.

Howley, P., & Dillon, E. (2012). Modelling the effect of farming attitudes on farm credit use: a case study from Ireland. *Agricultural Finance Review*, 72(3), 456-470.

Jenkins, S.P. (2017). Pareto models, top incomes and recent trends in UK income inequality. *Economica*, 84(334): 261-289.

Katchova, A. L. (2008). A comparison of the economic well-being of farm and nonfarm households. *American Journal of Agricultural Economics*, 90(3), 733-747.

Kerm, P.V. (2007). Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC. *Working Paper Series* 2007-01, Centre d'Etude de Population, de Pauvreté et de Politiques Socio-Economiques - International Network for Studies in Technology, Environment, Alternatives, Development

Kish, L. (1995). *Survey Sampling*. Chichester.

Kott, P. S. (1991). A Model-Based Look at Linear Regression with Survey Data. *The American Statistician* 45 (2): 107.

Lausev, Jelena. (2014). What has 20 years of public–private pay gap literature told us? Eastern European transitioning vs. developed economies. *Journal of Economic Surveys* 28.3: 516-550.

Lohr, Sharon L. 2000. *Sampling : Design and Analysis*. Nelson Education.

Matthews A. (2016). Focus on the distribution of direct payments. August 2016. Available at: capreform.eu (last accessed 20 June 2020).

Mendola, M. (2007). Agricultural Technology Adoption and Poverty Reduction: A Propensity-Score Matching Analysis for Rural Bangladesh. Food Policy, 32 (3): 372–93.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of political economy*, 66(4), 281-302.

Mishra, A., El-Osta, H., & Gillespie, J. M. (2009). Effect of agricultural policy on regional income inequality among farm households. *Journal of Policy Modeling*, 31(3), 325-340.

Mishra, A. K., & El-Osta, H. S. (2008). Effect of agricultural policy on succession decisions of farm households. *Review of Economics of the Household*, 6(3), 285-307.

Mishra, A. K., El-Osta, H. S., Morehart, M. J., Johnson, J. D., & Hopkins, J. W. (2002), Income, Wealth, And The Economic Well-Being Of Farm Households. *Agricultural Economics Reports*.

Mishra, A. K., & Sandretto, C. L. (2002). Stability of farm income and the role of nonfarm income in US agriculture. *Review of Agricultural Economics*, 24(1), 208-221.

Mishra, A., & Moss, C. B. (2008). Measurement of inequality by components of farm household consumption expenditures. *Applied Economics*, 40(10), 1241-1252.

Nguyen, B. T., Albrecht, J. W., Vroman, S. B., & Westbrook, M. D. (2007). A quantile regression decomposition of urban–rural inequality in Vietnam. *Journal of Development Economics*, 83(2), 466-490.

Pender, J., & Gebremedhin, B. (2006). Land management, crop production, and household income in the highlandpender of Tigray, Northern Ethiopia: An econometric analysis. *Strategies for sustainable land management in the East African highlands*, 107-139.

Phimister, E., Roberts, D., & Gilbert, A. (2004). The dynamics of farm incomes: Panel data analysis using the farm accounts survey. Journal of Agricultural Economics, 55(2), 197-220.

Rocchi, B., Marino, M., & Severini, S. (2020). Does an Income Gap between Farm and Nonfarm Households Still Exist? The Case of the European Union, *Applied Economic Perspectives and Policy*, https://doi.org/10.1002/aepp.13116

Rocchi, B., Stefani, G., Romano, D., & Landi, C. (2012). Are Italian farming households actually poorer than other non agricultural households? An empirical analysis, *Paper Prepared for Presentation at the 1st AIEAA Conference 'Towards a Sustainable Bio-Economy: Economic Issues and Policy Challenges.'*

Safari, M.A.M., Masseran, N. and Ibrahim, K. (2018). Optimal threshold for Pareto tail modelling in the presence of outliers. *Physica A - Statistical Mechanics and Its Applications*, 509: 169-180.

Safari, M.A.M., Masseran, N. and Ibrahim, K. (2019). On the identification of extreme outliers and dragon-kings mechanism in the upper tail of income distribution. *Journal of Applied Statistics*, 46(10): 1886-1902.

Schaffer, M., & Stillman, S. (2006). XTOVERID: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2, xthtaylor, Statistical Software Components S456779. Boston College Department of Economics. Revised Nov 2, 2011.

Severini, S., & Tantari, A. (2015). Which factors affect the distribution of direct payments among farmers in the EU Member States? *Empirica (Journal of European Economics)* 42(1): 25-48.

Stefani, G., Rocchi B., and Romano D. (2012). Does Agriculture Matter? Revisiting the Farm Income Problem in Italy. *Working Papers - Economics*, wp2012_18.rdf, Universita' degli Studi di Firenze, Dipartimento di Scienze per l'Economia e l'Impresa

Templ, M., Gussenbauer, J. and Filzmoser P. (2019). Evaluation of robust outlier detection methods for zero-inflated complex data. *Journal of Applied Statistics*, 47(7): 1144-1167.

United Nations (2012). *Statistics on rural development and agricultural household income. The*

Wye Group Handbook. Second Edition. Retrieved from http://www.fao.org/economic/ess/ess-capacity/wyegroup/it/. (last accessed January 2018).

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data MIT Press. Cambridge, MA, 108.