ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/320323281

Robustness Analysis of Organic Technology Adoption: Evidence from Northern Vietnamese Tea Production

Article *in* European Review of Agricultural Economics · May 2019

DOI: 10.1093/erae/jbz018

CITATIONS
2

READS
375

3 authors:

Nicolas Lampach
KU Leuven
35 PUBLICATIONS 41 CITATIONS

SEE PROFILE

Phu Nguyen-Van
University of Strasbourg
78 PUBLICATIONS 1,570 CITATIONS

SEE PROFILE

Nguyen To-The
Vietnam National University of Agriculture
18 PUBLICATIONS 34 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Euthority View project

Project Restructuring the tea industry towards organic tea production View project

All content following this page was uploaded by Nicolas Lampach on 09 October 2018.

The user has requested enhancement of the downloaded file.

Robustness Analysis of Organic Technology Adoption: Evidence from Northern Vietnamese Tea Production[§]

Nicolas Lampach[a*] Phu Nguyen-Van[b]

Nguyen To-The[c]

[a] *Centre for Legal Theory and Empirical Jurisprudence, KU Leuven (Belgium)*

[b] *BETA, CNRS, INRA & University of Strasbourg (France) and TIMAS, Thang Long University (Vietnam)*

[c] *Vietnam National University of Agriculture (Vietnam)*

Abstract

Increasing consumer awareness on sustainable and healthy food choices gave rise to a growing demand for organic tea in the past decades. Most of this demand is met by imports from developing countries. This article examines the main factors affecting the choice of farm households to adopt organic tea production in Northern Vietnam. We apply a logit model to survey data from 241 Vietnamese tea farming households. We assess the robustness of the results by addressing three important statistical issues: (i) regressor endogeneity, (ii) unobserved heterogeneity at farm level and (iii) missing values. The main results are chiefly robust and largely in line with the theory. We find that farm households with higher revenues, and located in rich natural and physical environments are significantly more inclined to adopt organic tea production. Furthermore, the analysis reveals that farm households who are consulted by extension agents and belong to a tea association increase the odds for the adoption of organic tea cultivation.

Keywords: Multiple imputation method, Organic farming, Regressor endogeneity, Tea production, Unobserved heterogeneity, Vietnam

JEL Classification: Q15; O33; Q18

[§] Nguyen To-The gratefully acknowledges the financial support from the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 502.01-2018.13. We thank Marco Buso, Patrizia Eleonora Ganci and Russel Neudorf for their helpful advice and proofreading. Help and support from colleagues of the economic department of the Vietnam National University of Agriculture for data collection are also acknowledged. We thank the editor of the European Review of Agricultural Economics and two anonymous reviewers for their concise and constructive comments during the review. All remaining errors are our own.

[*] *Corresponding author*: Nicolas Lampach. Address: Centre for Legal Theory and Empirical Jurisprudence, International House, 45 Tiensestraat, 3000 Leuven, Belgium; Tel: +32 16 37 76 08; E-mail: `nicolas.lampach@kuleuven.be`.

Robustness Analysis of Organic Technology Adoption: Evidence from Northern Vietnamese Tea Production

Increasing consumer awareness on sustainable and healthy food choices gave rise to a growing demand for organic tea in the past decades. Most of this demand is met by imports from developing countries. This article examines the main factors affecting the choice of farm households to adopt organic tea production in Northern Vietnam. We apply a logit model to survey data from 241 Vietnamese tea farming households. We assess the robustness of the results by addressing three important statistical issues: (i) regressor endogeneity, (ii) unobserved heterogeneity at farm level and (iii) missing values. The main results are chiefly robust and largely in line with the theory. We find that farm households with higher revenues, and located in rich natural and physical environments are significantly more inclined to adopt organic tea production. Furthermore, the analysis reveals that farm households who are consulted by extension agents and belong to a tea association increase the odds for the adoption of organic tea cultivation.

Keywords: Multiple imputation method, Organic farming, Regressor endogeneity, Tea production, Unobserved heterogeneity, Vietnam

JEL Classification: Q15; O33; Q18

1 Introduction

Increasing consumer awareness of food safety and quality has spurred the expansion of international trade in high-value food products over the last decades. Policymakers have recognized the potential of organic farming as a measure to ensure sustainability and meet the mounting demand of high quality food products. Organic farming[1] has received ample attention in most countries where food safety problems are present and gaps in the food supply of ecological friendly products exists. Since 1980, the development of organic farming–which prohibits the use of any synthetic agrochemicals–was mainly driven by committed farmers and consumers in the United States and Europe. Organic production has increased in developing countries due to the high demand of organic products in developed countries. Despite the positive trend, the overall rate of conversion to organic farming in developing countries has been observed increasing much slower than expected.

Tea is the most consumed manufactured beverage in the world and represents an important commodity in terms of labor and income for a large number of developing countries (FAO, 2015). In Vietnam, tea is one of the main export crops and the country is ranked as one of the top ten in the world in terms of tea production and exports (Tran, 2009). Between 2011 and 2021, Vietnamese tea production is predicted to increase from 115.696 to 148.101 tonnes, reflecting an annual average growth rate of 2.5 percent (FAO, 2012b). Despite this projected growth, Vietnam is facing difficulties in fostering trademarks and complying with food safety standards. The current poor quality and low level of product safety of Vietnamese tea are mainly attributable to monoculture farming

[1] There exist many definitions of organic farming. A more general form can be found in the Codex Alimentarius Commission created in 1963 by the Food and Agriculture Organization of the United Nations (FAO) and the World Health Organization (WHO) referring to "*organic farming involves holistic production management systems (crops and livestock) emphasizing the use of management practices in preference to the use of off farm inputs*".

and the misuse of chemicals (i.e. fertilizers and pesticides) (Banerjee, 1981; Thang et al., 2015).

In August 2005, the Vietnamese Tea Association (VTA) recommended a package of measures (e.g. standards, certificates, predetermined level of industrial hygiene, quality control system) to improve the quality of tea. After this point, the Ministry of Agriculture and Rural Development in Vietnam started to encourage a shift to organic farming by setting organic tea production as a key priority into Vietnam's National Policy Framework. To take advantage of this opportunity and reap the benefits of organic tea farming several obstacles need to be overcome. First, certification programs can increase the visibility and marketability of organic products however those schemes are cost-intensive for farmers and the standard obligations are often administratively burdensome (Van Bac et al., 2017). The legal transition to organic farming lasts two to three years during which products cannot be sold on the market as organic. Initial loss of yields during this transition period and the state of ecosystem degradation from previous management practices are common constraints that can only be offset if sufficient financial support is given to farmers (Scialabba, 2000).

The organic market in Vietnam is very small and there exists only a limited number of Vietnamese tea products certified as organic. Nowadays, larger amount of land devoted to tea production in Vietnam is being converted to organic farms (Tran, 2009). In 2001, only 2 hectares of organic land were managed by 38 farms in Vietnam representing 0.003 percent of total agricultural area. In 2015 the certified organic area was 37.490 hectares representing 0.36 percent of the total agricultural area of Vietnam (Hsieh, 2005, Willer and Kichler, 2015). Furthermore, a switch from conventional to organic production might be beneficial for both producers and society as it reduces the health and environmental impacts of pesticide usage. Understanding the determinants that affect the adoption of organic farming practices might be crucial for Vietnamese agricultural policy in order to establish effective measures and foster export growth of high-value food products.

The bias associated with endogeneity has received widespread attention from scholars in recent years, but surprisingly little attention has been given to other statistical problems caused by unobserved heterogeneity and the occurrence of missing values in survey or longitudinal data. Unobserved heterogeneity can be related to causing non-observable differences and variations among observations. It might be plausible to argue that the adoption decision can be substantially influenced by farmers' motivation, attitudes or personal characteristics which are simply not observed by econometricians. The problem of missing values is ubiquitous in survey data collection efforts as respondents might have left out answers to questions for several reasons. In particular, the presence of unobserved heterogeneity and missing values can lead to inefficient estimation and thus give inconsistent results. Hence, it becomes necessary to undertake a rigorous robustness analysis to draw credible inference from the estimated results.

This study aims to determine the main characteristics influencing the adoption of organic farming in the Vietnamese tea sector. We use cross-sectional data of 241 tea farming households collected by ourselves through a survey questionnaire in 2013. Logit model estimation is used to analyse the relationship between the key determinants and the adoption of organic tea production at farm-household level. The robustness of the results are assessed from three statistical perspectives: (i) regressor endogeneity, (ii) unobserved heterogeneity at farm level and (iii) missing data from the survey.

In the agricultural adoption literature, extensive studies have found significant effects of gender, education, concern for the environment, trust in government and market, and information services on the adoption of alternative agricultural practices, likewise traditional, organic or agro-forestry systems (Baumgart-Getz et al., 2012; Burton et al., 2003; Darnhofer et al., 2005; Fairweather and Campbell, 2003; Flaten et al., 2005; Haugen and Brandth, 1994; Karki et al., 2012; Koesling et al., 2008; Läpple, 2010; Läpple and Kelley, 2015). Of utmost interest are also the articles by Jansen (2000) and Padel (2001) reviewing a large number of studies of organic farming related to the adoption model. Scholars primarily apply non-linear models to estimate the main factors of the adoption of alternative agricultural practices where the dependent variable is dichotomous.[2]

This article makes a contribution to the extensive adoption literature in the field of organic farming by focusing on a novel survey database and providing support for policy actions by promoting organic farming. Secondly, this article contributes to the methodology literature by assessing more rigorously the robustness of the results while considering three alternative specifications in the empirical analysis.

The remaining of this article is organized as follows. Section 2 motivates the hypotheses from the adoption theory. Section 3 presents the data used in the analysis. Section 4 introduces the empirical framework and outlines the main findings. Section 5 discusses four different alternative specifications to deliver robust results and section 6 provides policy prescriptions. Section 7 concludes and gives an outlook on the Vietnamese agricultural policy and organic tea sector.

[2] We acknowledge that a continuous variable–share of land under the new crop or the duration of the adoption–indicating to what extent a new crop is adopted might be more informative. Unfortunately, we neither have accurate indicators about the percentage of land cultivated by organic farming nor the duration of the adoption in our database. Related studies on the share denomination of organic and conventional farming are amongst others, Arslan et al. (2014), Croppenstedt et al. (2003), Khaledi et al. (2010) and Kuminoff et al. (2005).

2 Motivations

There has been a constant effort to explain farmers' adoption choice of particular agricultural technologies. Researchers have investigated a wide array of possible explanations to determine farmers' adoption behavior. We chose a selection of studies to particularly motivate four relevant hypotheses in order to gain an improved understanding of the aspects of organic adoption in general.

A key determinant in the adoption literature is farmers' income. There is extensive literature examining why poor households in developing countries tend to be less inclined to adopt new agricultural technologies. Explanations that have received attention in policy discussion and academic research are the availability of inputs, uncertainty of profitability, credit and insurance constraints (Alem and Broussard, 2018, Feder et al., 1985). The studies by Dey et al. (2010), Negatu and Parikh (1999) and Udensi et al. (2011) demonstrate that farm households with higher incomes are more likely to adopt innovations compared to those with lower incomes. Two main reasons for this result are farmers' positive perception for the marketability of the modern crop and greater financial feasibility for investigating new technologies. Moreover, Tran (2015) highlights that Vietnamese small-scale farmers with higher incomes can afford higher investment costs and therefore tend to be more likely to respond effectively to climate change than farmers with lower incomes.

Other critical factors emphasized by the adoption literature are the acquisition of new information from associations or through extension services. Various studies have shown that farming associations enhance the interaction among farmers (Abdoulaye et al., 2014, Adebayo and Oladele, 2013, Ojiako et al., 2007, Owusu et al., 2013, Versteeg and

Koudokpon, 1993). Indeed, agricultural associations play a crucial role in supporting farmers to work together through collective participation in markets and value chains. Membership in a farmers' group is considered as an important access to information. It elucidates a network in which farmers share experiences and can learn from each other through direct collaborations. Another crucial point is that farming associations may reduce barriers and costs for new farmers by providing better access to credits or the collective use of agricultural machinery (Banson et al., 2018). Empirical evidence from developing countries suggests that cooperative membership has a strong positive and significant impact on agricultural technology adoption (Abebaw and Haile, 2013; Adesina and Baidu-Forson, 1995; Ahmed and Mesfin, 2017; Ma, 2016; Wollni and Zeller, 2007; Wossen et al., 2017).

Another form of governmental support for technological adoption is agricultural extension services. Effective extension programs can bridge the gap between new discoveries in the laboratory and practical changes carried out by farmers. Especially, extension agents can inform farmers about new cropping techniques, high yield varieties but also about their managerial skills by shifting toward more efficient production methods (Birkhaeuser et al., 1991). The findings from Abdoulaye et al. (2014), Adesina and Zinnah (1993), Akinola et al. (2010), Ali and Abdulai (2010), Chirwa (2005), Owusu et al. (2013) and Shiferaw and Holden (1998) reveal that farmers who are informed by extension agents tend to adopt more likely the agricultural innovations than those who are not consulted. Bryan et al. (2009) report that the effect of extension services on the adaptation to climate change appears to be larger for the group of poor farm households than for those with higher income.

The probability of adopting new modern technology may also depend on the natural and physical environment of cultivation. Higher soil quality, better water availability and efficient irrigation systems increase the expected utility of income from modern production and thus elevate the likelihood of adopting new technology. Caswell and Zilberman (1986) study in a theoretical framework the effect of land quality and well depth on farmers' adoption of modern irrigation technologies and present the conditions under which farmers are more likely to adopt the new agricultural technology. Other formal models suggest that farms which are located in more favorable environments will be more likely to adopt new production method (Hiebert, 1974; Nelson and Phelps, 1966; Welch, 1970). However, only a scarce number of empirical studies consider micro-level variables (e.g. soil characteristics, field slope, temperature, field gradient, water-holding capacity) in the evaluation of farmers' choice of technology. Green et al. (1996) assess the effect of economic, environmental and institutional variables on irrigation technology adoption. Their empirical findings highlight that agronomic and physical characteristics influence positively and significantly farmer's adoption behavior. It has been also underlined that environmental variables appear to matter more than economic factors. Based on the insightful studies from the adoption literature, our four hypotheses can be summarized as following:

- *Hypothesis 1: Farm households with larger revenues are more likely to adopt organic tea production*
- *Hypothesis 2: Farm households being part of a tea association are more likely to adopt organic tea production*
- *Hypothesis 3: Farm households who have been consulted by extension agents are more likely to adopt organic tea production*
- *Hypothesis 4: Farm households located in a rich natural and physical environment are more likely to adopt organic tea production*

Now that we have identified our four hypotheses, we explain the data and variables used for the empirical analysis in the following section.

3 Data

A survey[3] was carried out from January to May 2013 by our team in three provinces located in Northern Vietnam, namely Tuyen-Quang, Phu-Tho and Thai-Nguyen (See Figure 1). The tea cultivation in Vietnam is primarily concentrated in two regions. The first is comprised of the three aforementioned provinces representing approximately 60% of the total area, and the second, the Central Highlands, representing 20% in 2009 according to the Vietnamese tea association. Nine representative communes[4] of the three tea-producing provinces were chosen for the survey. The selected communes are representative of topographical and climate conditions in tea production areas in the three provinces in Northern Vietnam. Participants were randomly selected from a list of farm households engaged in tea production.

– – – [Figure 1 here] – – –

We asked the participants to provide information about their tea production of the previous year. Additionally, we organized face to face interviews with the head of household. The average duration of the questionnaire lasted 1 hour and 13 minutes with a maximum of 2 hours. Quantitative and qualitative information have been collected for a total sample of 241 Vietnamese farm households. Originally, a random sample of 250

[3]Note that local enumerators conduct the survey. They were prior trained by staff members while receiving general instructions and exercises.

[4]Van Linh (Thanh Ba district), Ngoc Dong (Yen Lap district), My Bang (Yen Son district), Phuc Triu (Thai Nguyen district), Hoang Nong and La Bang (Dai Tu district), Ba Xuyen, Binh Son and Tan Quang (Song Cong district). Table 5 in Supplementary Materials lists the number of population and the farm households by tea province.

farm households from the overall population (i.e 11006 listed farms) have been selected, but unfortunately 9 farm households did not participate owing to time constraints.

Although our sample constitutes 241 potential adopters, this can be viewed as a moderate sample size. This sample is obtained from a random sampling procedure, which provides similar figures compared to the values related to the total Vietnamese tea sector reported by the General Statistical Office (GSO). According to GSO (2011), the productivity of Vietnamese tea constitutes about 5-18 tons/ha during the period 1961-2011 whereas our sample presents an average productivity of 6.72 tons/ha and a standard deviation of 5.75 tons/ha (the distribution range is comprised between 0.10 and 23.33 tons/ha). The obtained sample is representative of the population of tea producers in Northern Vietnam (i.e three provinces).

A summary of the data and variables used in the empirical analysis are presented in Table 1. Recall that the dependent variable is CHOICE, a binary variable indicating whether the farm household adopts organic tea cultivation. More precisely, this variable equals one if the farm opts to produce organic tea, otherwise equals zero (the farm produces conventional tea). The set of explanatory variables includes REVENUE, LAND, EXPERIENCE, HHSIZE, HEDUC, MINORITY, GENDER, TASSO, EXTENSION, and province dummies (for provinces Tuyen-Quang, Phu-Tho, and Thai-Nguyen, the last dummy being the reference). Except for descriptive statistics reported in Table 1 corresponding to the levels of household's income, land area, and labor, we use the logarithm values of REVENUE, LAND, and LABOR in the estimations to reduce the heterogeneity and presence of possible outliers in the data.

$$- - - \Big[\text{Table 1 here}\Big] - - -$$

The variable REVENUE is measured in million VND (21,148 VND is equivalent to

1$ indicated by the World Bank). The average tea revenue is about 65.7 million VND (3.106 USD) per farmer, with a standard deviation of 67.1 million VND (3.173 USD). The differences between the group of adopters and non-adopters related to household income is significantly different at 5% level (p_{value}=0.014) applying Kolmogorov-Smirnov test.[5] One could object that the variable revenue might not be appropriate to use in the model. But, as we do not have any other source of data for these groups of farmers, we were left with no other choice. Instruments for revenue were sometimes not available to remove the endogenous problem of farm revenue per cultivated land area (Adesina and Chianu, 2002; Dey et al., 2010; Negatu and Parikh, 1999; Udensi et al., 2011 and Zeller et al., 1998, except Kan et al. (2006)).

The variable LAND represents the total farm size of the farm household measured in hectares. The average farm land is about 0.58 hectares per household. As shown in the summary statistics the variable takes on a wide range of values. For this reason, we use the variable in logarithm form to reduce the variability in the subsequent estimation.

The average experience of farmers is 29.7 years with a standard deviation of 13.8. The duration of farmers' experience may effect positively or negatively the adoption choice. Young farmers have been found to be more willing to bear risk, more knowledgeable and more likely to adopt new practices due to longer planning horizons. Older farmers are less likely to adopt new practices as they feel confident in cultivating tea in their old ways and method. On the other hand, old farmers may have more experience, resources, or authority that may endow them with more capabilities to adopt new practices (Abdoulaye et al., 2014; Abebe et al., 2013; Nguyen-Van et al., 2004).

[5] In Supplementary Materials, Table 6 summarizes the differences between both groups adopters and non-adopters related to all covariates. This test generally shows the existence of differences between these two groups, in terms of revenue, cultivation surface, household size, membership of a tea association, and provinces.

The average number of members in a household size is 4.3, with a standard deviation of 1.2. In the adoption literature, the variable household size has been identified to affect either positively or negatively the decision to adopt a new modern technology (Abdoulaye et al., 2014; Kafle and Shah, 2012; Kebede et al., 1990; Shiferaw and Holden, 1998; Staal et al., 2002; Udensi et al., 2011; Zeller et al., 1998).

We include dummies corresponding to household characteristics, and likewise for higher education, minority, gender, tea association (TASSO) and extension (EXTENSION). In our data, 80 households have a higher education, 26 households are belonging to a minority ethnic group, 79 households are members of a tea association and 173 households have been contacted by extension agents.

HEDUC is a proxy for the educational level of the head of household and is measured as a dummy variable which takes the value one if the farmer benefited from formal school education and zero otherwise. Studies have shown that farmers with higher education tend to be more likely to adopt new innovations (Abebe et al., 2013; Abdoulaye et al., 2014; Kafle and Shah, 2012; Ouma and De Groote, 2011; Padel, 2001; Storstad and Bjørkhaug, 2003). GENDER is a dummy variable and equals one if the head of household is male and zero otherwise. Scholars argue that women are generally discriminated in terms of access to external inputs and are less likely to adopt new modern technology (Abdoulaye et al., 2014; Adesina, 1996; Haugen and Brandth, 1994; Jansen, 2000).

A large majority of farm households (57%) are located in the mountainous midland province of Thai-Nguyen which constitutes one of the largest tea producing areas in Vietnam. The climate conditions (temperature and rainfall) make the province favorable for agricultural development. In spite of these conditions, arable land only constitutes 12.4% of the total land area and is exposed to droughts and floods which make harvesting more difficult. Another 30% of farm households are situated in the province of Tuyen-Quang and the minority (13%) are located in the province of Phu-Tho. The latter represents one of the poorest areas of Vietnam and households depend largely on tea cultivation. The province Phu-Tho is located in a subtropical monsoon region along the Red River Delta. With an annual rainfall of 1680mm, average minimum and maximum temperatures of 14.0 $°C$ and 32.7 $°C$, and ferralitic soil, the province offers excellent pedoclimatic conditions to grow tea trees. Phu-Tho is one of the most important provinces for tea production and tea plays a key role in the development of local agriculture (Chuc et al., 2006; FAO, 2012a; Ninh et al., 2017; Pandolfi et al., 2009). In the past, the establishment of the tea processing industry and the incorporation of research and development centers helped to promote the production of tea. In 1918 the French established the first tea plantation in Phu-Tho province and helped to set up the Agricultural and Forestry Research center seeking to investigate effective growing conditions and methods (Wenner, 2001). In the last few years, French development aid programs financed new plantations in the province of Pho-Tho between 2001 and 2004 to enhance the economic development of rural population and improve their living conditions French development agency, 2004.

4 Preliminary empirical results

To analyse the adoption of organic tea production, we employ the following logit model (see Amemiya, 1981 and Maddala, 1991):

$$y_i^* = x_i\beta + \varepsilon_i, \tag{1}$$

where y_i^* is a latent variable which represents the net utility corresponding to organic tea production compared to traditional (or conventional) tea production. We observe the decision of tea producer i, which is a binary variable, i.e. $y_i = 1$ if producer i adopts organic tea production and $y_i = 0$ if (s)he selects the traditional production. This corresponds to variable CHOICE in the data. The adoption decision depends on producers' characteristics and other control variables, represented by x_i, and an unobserved random factor ε_i. We adopt the logit specification with the following probability

$$Pr(y_i = 1 \mid x_i; \beta) = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}, \tag{2}$$

and the corresponding log-likelihood function

$$\ln L = \sum_{i=1}^{n} \left\{ \mathbf{I}(y_i = 1)\ \ln Pr(y_i = 1 \mid x_i; \beta) + \mathbf{I}(y_i = 0) \ln \left[1 - Pr(y_i = 1 \mid x_i; \beta)\right] \right\}, \tag{3}$$

where $\mathbf{I}(.)$ is the indicator function of producer's choice.

Table 2 reports estimation results (coefficients and marginal effects) of the logit model concerning the choice of organic tea production.[6] Estimation results for the logit model show that farm households having higher revenues, belonging to a tea association, being consulted by extension agents and located in the Phu-Tho province are significantly more likely to adopt organic tea cultivation.

In particular, for each additional unit of (log) revenue, the farm is 8.8% more likely to produce organic tea in comparison with those that cultivate conventional tea. This result corroborates with *Hypothesis 1* and is consistent with the findings of Dey et al. (2010),

[6]It is well known that the marginal effect of explanatory variable x_k on adoption probability is given by $\partial p / \partial x_k = p(1-p)\beta_k$ with $p \equiv Pr(y_i = 1 \mid x_i; \beta)$. Note that the marginal effect of a continuous explanatory variable represents changes in the dependent variable when the explanatory variable increases by one unit. The marginal effect of a categorical explanatory variable is differently computed, i.e. it corresponds to changes in the dependent variable when the latter shifts from zero to one.

Negatu and Parikh (1999) and Udensi et al. (2011). Economic factors play an important role in the adoption decision of new technologies (Adesina and Chianu, 2002). It is argued that wealthier farm households tend to exhibit more risk-seeking behavior as they may be able to compensate for potential losses. Our proxy variable LN REVENUE partly includes the information of the tea price. One could object in terms of estimation that the proxy revenue as an exogenous variable might not be appropriate and that a reverse effect can exist as the choice of organic farming may have an impact on the household's income. The next section will be devoted to the possible endogeneity of LN REVENUE as well as EXTENSION (contact with extension agents).

$$- - - \left[\text{Table 2 here}\right] - - -$$

Several authors have used recursive econometric models to explain the adoption of agricultural technology and related income effects (Zeller et al., 1998). A similar framework is applied in this article. We define the adoption of organic tea and the resulting revenue generation as a sequential decision making process whereby previous cropping decisions predetermine farm revenue. A tea association member is significant and 18.6% more likely to produce organic tea than a non-member. This result is consistent with *Hypothesis 2* and in line with the literature revealing a positive relationship between the membership of a tea association and the adoption of organic tea cultivation. Farmers who are part of a tea association are more likely to adopt the production of organic tea (Saka et al., 2005; Owusu et al., 2013 and Adebayo and Oladele, 2013). Agricultural development agencies have high rates of success when they work with farmers' groups or associations (Versteeg and Koudokpon, 1993).

Farm households consulted by extension agents are significantly and 8.6% more likely to produce organic tea than those who are not consulted. In line with *Hypothesis 3*, the extension activities reflect the efficiency of the agricultural extension system in recent years. This finding is to some extent in line with the results of Abdoulaye et al. (2014), Adesina and Zinnah (1993), Ali and Abdulai (2010), Ojiako et al. (2007) and Owusu et al. (2013). Meanwhile, Adebayo and Oladele (2013) also show that farmers who have been contacted by extension agents are less likely to practice organic farming techniques. This variable can be potentially endogenous because of a reverse effect: the adoption of organic tea production can lead to more frequent contact with extension agents as farmers would require the need for more information in this case than with traditional/conventional tea. We will address the endogeneity of this variable in the next section.

In line with *Hypothesis 4*, our estimation results demonstrate that farm households situated in the province of Phu-Tho are significantly and 19% more likely to cultivate organic tea compared to farm households located in Thai-Nguyen (which serves as the reference category). The marginal effect of the province is of substantially large size and might be influenced by other relevant factors.[7] In the last years, several initiatives gave rise to the production of organic products in the province of Phu-Tho. For instance, a seven-year organic farming project during 2005-2012 was launched by the Danish Government in assistance with the Vietnam Farmer Union. The main objective was to increase awareness and knowledge of organic farming and provide assistance in the production and marketing of products (Dam, 2016).

Several authors mentioned that both education and experience play a significant role in the adoption decision (Abebe et al., 2013; Abdoulaye et al., 2014; Adesina and Chianu, 2002; Ayuk, 1997; Kafle and Shah, 2012; Kebede et al., 1990; Langyintuo and Mungoma, 2008; Ouma and De Groote, 2011; Saka et al., 2005; Strauss et al., 1991; Schultz, 1975).

[7] We acknowledge that we are not able to disentangle the effect of climate conditions and soil from other factors like road rehabilitation, transport access, development program aid, local market, research centers or industrial parks.

However, these variables were not found to be significant in our preliminary estimation of the data. Furthermore, the variable gender has been considered to be important but we could not find any significant effect in our model.

5 Robustness

Different alternative specifications and data treatments can be employed to deliver robust results. We consider four possible issues. The first one is the inclusion of unobserved heterogeneity into the choice model as we think that a farmer's decision can be linked to some motivation and characteristics that econometricians are not able to observe. This problem can give inefficient estimation (i.e. the estimator does not have the smallest variance). The second case deals with the problem of endogenous regressors that can bias estimation results. The third issue is related to missing values in the survey as it can alter the quality of the regressions (i.e. the estimator might be inconsistent). Finally, the last issue is the goodness-of-fit of the selected model.

5.1 Farmer's unobserved heterogeneity

To obtain a more general specification, we allow for the presence of a random effect (or producer's unobserved heterogeneity).[8] In this case, the probability of adopting organic tea production becomes

$$Pr(y_i = 1 \mid x_i, \mu_i; \beta, \sigma) = \frac{\exp(x_i'\beta + \sigma\mu_i)}{1 + \exp(x_i'\beta + \sigma\mu_i)}, \tag{4}$$

[8] A more general framework than our specification here is the mixed logit model following the ample precedent of McFadden and Train (2000) where the model parameters β are assumed to be random, for example $\beta = \bar{\beta} + \zeta$ where ζ is distributed following some standard density $f(\zeta)$. This alternative modelling is left for further research.

and the probability of choosing the conventional tea production is similarly defined. The heterogeneity term μ_i is assumed to be mutually independent of x_i, and is assumed to be standard normally distributed. The assumption about the i.i.d. distribution for μ_i helps us to easily compute the maximum-likelihood estimation. On the contrary, if individual heterogeneity is not i.i.d., the integration should be made over the whole sample, which is not really tractable. The log-likelihood function of the model is

$$\begin{aligned} \ln L &= \sum_{i=1}^{n} \left\{ \mathbf{I}(y_i = 1)\ \ln \int Pr(y_i = 1 \mid x_i, \mu_i; \beta, \sigma)\varphi(\mu_i)\mathrm{d}\mu_i \right. \\ &\quad \left. + \mathbf{I}(y_i = 0) \ln \left[1 - \int Pr(y_i = 1 \mid x_i, \mu_i; \beta, \sigma)\varphi(\mu_i)\mathrm{d}\mu_i \right] \right\}, \end{aligned} \tag{5}$$

where individual heterogeneity should be integrated out following its probability distribution $\varphi(\mu_i)$ (i.e. standard normal). For this purpose, we compute the integration by taking the average over a number R of pseudo random draws for μ_i^r.[9] Estimates for β and the additional parameter σ are obtained by maximum likelihood.

We compare two nested models, without and with unobserved heterogeneity by calculating the likelihood ratio test. As shown in Table 2, the test statistic is very low, 0.012, which is much lower than the 5% critical value of the $\chi^2(1)$ distribution, i.e. 3.84. Therefore, the model without unobserved heterogeneity (as presented in Section 4) is preferred.

5.2 Endogenous regressors

It should be noted that some explanatory variables could be potentially endogenous. In particular, the presence of omitted variables and the possible reverse causality can

[9] The expression of this estimator is given by $(1/R)\sum_{r=1}^{R} Pr(y_i = 1 \mid x_i, \mu_i^r; \beta, \sigma)$. In estimation, we define $R = 100$, given that other values ($R = 50, 200, 300$) do not change the results. Further details can be found in McFadden and Train (2000).

induce the endogeneity of LN REVENUE and EXTENSION. Omitted factors can cover other household's characteristics, production technology, and policy variables that are not observed in the data. The reverse effect stems from the fact that the adoption of organic tea production can help households to improve their income on the one hand, and on the other hand can motivate their participation in agricultural extension programs set up by the government. Moreover, the non-adoption of organic production could also motivate the participation of farm households in extension programs. These arguments suggest that the potential endogeneity of both variables LN REVENUE and EXTENSION should be appropriately accounted for in order to identify their causal effects.[10] The question of regressor endogeneity in nonlinear models (such as the logit model here) is a relatively recent advance in econometrics. We test the endogeneity of LN REVENUE and EXTENSION by implementing the following two-step procedure of Wooldridge (2014).

1. First, we make an OLS regression for LN REVENUE on all of the explanatory variables above (except EXTENSION) and a set of excluded instruments. For EXTENSION, which is a binary variable, we make a probit regression on the same set of explanatory variables (except LN REVENUE) and a set of excluded instruments. We observe that the linear regression for LN REVENUE looks like a production function whereas the probit regression for EXTENSION corresponds to the usual participation equation. We use the same set of excluded instruments for both of them, i.e. log of labor, use of chemical fertilizers, and use of organic fertilizers. The step allows us to compute the residuals for the first regression, $\hat{u}_i$, and the general-

[10] It is reasonable to assume that farm size is an exogenous variable in our model. Indeed, in the context of Vietnam, it is very difficult to modify farm size because of two reasons. On the one hand, tea tree only starts giving a significant production if these are at least 5 years old. Hence, the cultivation surface appears to be unlikely an adjustment factor in the short run. On the other hand, the market for land use rights is highly controlled by the government in Vietnam so that transaction is almost inexistant (land belongs to the ownership of all the people and is managed by the State, private ownership of land is not permitted in Vietnam).

ized residuals for the second regression, $\hat{v}_i = w_i \lambda(z_i'\hat{\gamma}) - (1 - w_i)\lambda(-z_i'\hat{\gamma})$ where w_i is the indicator for extension participation (EXTENSION = 1) and λ is the inverse Mills ratio ($\lambda(.) = \phi(.)/\Phi(.)$, ϕ and Φ being respectively the density and cumulative probability of the standard normal distribution).

2. Secondly, we perform the usual logit regression (with and without unobserved heterogeity) as described above with two additional explanatory variables, $\hat{u}_i$ and $\hat{v}_i$, computed in the previous step. The endogeneity of LN REVENUE and EXTENSION is therefore tested by using robust Wald test for the null hypothesis that the coefficients of $\hat{u}_i$ and $\hat{v}_i$ are jointly zero. The test is called '*robust*' because it is based on the robust variance-covariance matrix. The test statistic corresponds to a $\chi^2(2)$ distribution.

Table 2 reports that the robust Wald test statistic is 1.86 and that the corresponding p-value is 0.39. Thus, we cannot reject the null hypothesis of absence of regressor endogeneity and conclude that LN REVENUE and EXTENSION are exogenous. Again, this finding confirms the model presented in the previous section.

5.3 Multiple imputation

Omitting missing values in empirical analysis may lead to biased and inconsistent parameter estimates. Indeed, Rubin (1976) pointed out that inference can remain correct if the missing data are missing at random. However, this condition is not always hold, i.e. inference is then conditional on the pattern of missing data. In other words, failing to deal with missing data might cause substantial parameter bias and influence the efficiency and explanatory power of the results (Cheema, 2014; Rose and Fraser, 2008). The problem of missing data is ubiquitous in social research and has received widespread attention in a large number of research disciplines Cheema, 2014; Ghosh, 2011, Gómez-Carracedo et al., 2014, Kalaycioglu et al., 2016; Li et al., 2015; Moss and Mishra, 2011; Rezvan et al., 2015). The current literature discusses many imputation methods, ranging from simple mean imputation –where missing values of a variable are replaced by the mean of non-missing data– to more complex parametric and semi-parametric imputation methods. The latter involves the specification of a model for the missing values given the observed data and draws imputed values from the posterior predictive distribution that can be different for each missing value (Heitjan and Little, 1991, Morris et al., 2014; Schenker and Taylor, 1996).

Multiple imputation methodology can be a powerful tool if the measures in the imputation model are associated with the missing values to produce less biased estimators in the presence of missing data. One major concern is whether the missing data is missing completely at random. In other words, missingness does not depend on the variables in the data set (Little, 1988). Multiple imputation analysis needs to be conducted very carefully as consistent results can only be obtained when the data is not missing completely at random (MCAR). MCAR means that there is no relationship between the data point and the missing data. Thus, MCAR assumes that the probability of the observation being missing does not depend on observed and unobserved measurements. Deleting missing values would only be desirable if the MCAR assumption holds. Therefore, we employ Little MCAR test to assess whether the missing values in our database are missing completely at random. MCAR test is statistically significant at 1% level (p-value< 0.001) and rejects the null hypothesis that the values are missing completely at random.

Missing values in our survey data accounts for 22%. Survey respondents in our sample might have left items blank on the questionnaire due to either lack of understanding, knowledge or unwillingness to answer the question. We apply the predictive mean matching (PMM) method which imputes a value randomly from a set of nearest observed value whose predicted values are closest to the predicted value from a regression model (Horton and Kleinman, 2007). In other words, a farm household with a missing value in a variable, it is replaced by the observed value of the farm household with the nearest predicted value. This method relaxes some of the parametric assumptions and is widely known to improve the robustness of inference with missing data to misspecification of the imputation model and ensures that the imputed values are plausible even if the normality assumption is violated (Horton et al., 2003; Morris et al., 2014; Vink et al., 2014). Supplementary Materials provides a tutorial on the application of multiple imputation method and depicts the distribution of observed and imputed data.

$$- - - \Big[\text{Table 3 here}\Big] - - -$$

We compute information criteria (AIC and BIC) to compare different models (with and without unobserved heterogeneity) estimated with either original or imputed data. The computed values are reported in Table 3. The results confirm that the model without unobserved heterogeneity and estimated with original data (i.e. the model presented in Section 4) is the best among the competing models as it corresponds to the lowest values of AIC and BIC.

5.4 Goodness of fit

We assess the goodness-of-fit of the selected logit model. Firstly, as shown in Table 2, the adjusted McFadden's pseudo R^2 is 0.327, which roughly means that the whole set of explanatory variables can explain about 33% of variation in the household's choice. This test is also consistent with the test for model significance, which rejects the hypothesis that all the coefficients of the model are jointly zero (test statistic $\chi^2(11) = 33.26$ with p-value $= 0.001$). For comparison purposes, the adjusted pseudo R^2 is 0.322 and 0.317 for the logit model with unobserved individual heterogeneity and the logit model with regressor endogeneity, respectively. This result shows that the logit model is the best fit for the original data. Using the imputation data, these figures are respectively 0.216 (logit), 0.215 (logit with unobserved heterogeneity), and 0.236 (regressor endogeneity), leading to the conclusion that the model with endogenous regressors is the best. However, it should be noted that the adjusted McFadden's pseudo R^2 cannot help compare models estimated over different data (i.e. models using original data versus models using imputation data). Therefore, for comparing models using these data, we can use the information criteria as in Section 5.3 above and the predictive score below.

$$- - - \Big[\text{Table 4 here}\Big] - - -$$

Additionally, we calculate the ratio of correctly predicted observations (including both adoption and non-adoption of organic tea production) over the total number of observations (named as a 'predictive score').[11] Table 4 provides the numbers of predictions in comparison with real observations and the predictive score for each of the three alternative specifications (logit, logit with unobserved heterogeneity, logit with endogenous regressors) using the original and the imputation data. We observe that the predictive score of the logit model and the logit model with unobserved heterogeneity are approximately equivalent (their values are very close: 0.872 and 0.896 for original data, 0.805 and 0.797 for imputation data) and are both much higher than that of the model with regressor

[11] We admit that this measure does not allow us to report the accuracy of prediction across separate components of the samples (adoption or non-adoption) because of the aggregation over the two components. Another measure, as proposed by an anonymous reviewer for the Bayesian framework, is the product between the two ratios of correct predictions corresponding to adoption and non-adoption separately. However, its advantage and weakness over our measure is unclear. This issue is left for future exploration.

endogeneity. This result underlines the better performance of the previous two models over the last one, either with original or imputation data. This finding is also consistent with the exogeneity test performed in Section 5.2 which concludes that LN REVENUE and EXTENSION are both exogenous.

6 Policy prescriptions

Results presented in Section 5 support the superiority of the simple logit model over alternative specifications. The analysis in Section 4 highlights the effects of household revenue, the presence of extension activity, and the membership of a tea association on the probability of organic tea adoption. In this section, we provide a deeper insight on the adoption probability in relation with these variables.[12]

Plotted in Figure 2 is the relation between adoption probability and household revenue using the sample means of other explanatory variables. This relation represents our prediction about the adoption of organic tea cultivation when household revenue increases. We observe that the logit model and the logit model with unobserved individual heterogeneity give a very similar relation while the model with endogenous regressors overestimates the adoption behavior (hence, an underestimation of the non-adoption) compared to the two previous models. This overestimation was already shown in Table 4: compared to two other models, the logit model provides a higher number of correct predictions for organic tea and a lower number correct predictions for conventional tea. This corroborates the conclusion about the supremacy of either the logit model or the logit model with unobserved heterogeneity over the model with endogenous regressors.

[12]We are thankful to an anonymous referee for suggesting this analysis. Similar presentation was included in previous literature but in another context, see Hattam et al. (2012), Holloway et al. (2014), among others.

$---\left[\text{Figure 2 here}\right]---$

Figure 2 illustrates the idea that it is hard to produce organic tea for low revenue households. An increase in revenue is an incentive for adopting organic cultivation.[13] Consequently, compared to rich households, poor households are less likely to adopt organic farming. This finding implies that a good policy should support these farmers to alleviate the costs of new agricultural technology. New policy measures in form of subsidies need to be established to encourage the conversion of poorer households to organic farming. The government could provide financial support to them in order to compensate the initial loss of yields and expenses in the transition period.

We show in Figure 3 the heterogeneity which can be observed in the adoption behavior. The first heterogeneity is geographic: among the three provinces, farmers residing in Phu-Tho province (Figure 3b) have the highest chance to adopt organic production, other things being equal. Two other heterogeneities are related to extension activity and participation to a tea association. The curves displayed in Figure 3 correspond to different combinations between them. Recall that the increasing pattern represents the positive effect of revenue on the adoption probability. However, the effect is stronger for farmers who are contacted by extension agents and are members of a tea association. This result underlines the role of the extension system and tea associations. If policy makers seek to promote organic farming, they can rely on these two channels, for example by intensifying existing extension programs, designing new extension activities, and fostering the development of tea associations and networks. In the context of Vietnam, we think that this prescription is within reach as the government has set up and continues to support several extension programs. Furthermore, Vietnam can also improve existing professional

[13] We can have a causal interpretation here as variable LN REVENUE is exogenous following the analysis provided in Section 5.2.

networks (although all of them are under the control of the Communist Party of Vietnam) in order to attract farmer participation.

$- - -\Big[$Figure 3 here$\Big] - --$

7 Conclusions

This article explores the main determinants of adopting organic tea production in Northern Vietnam. We employ an empirical analysis using a logit model. The robustness of the results are assessed for three important statistical questions (regressor endogeneity, unobserved heterogeneity and missing data). The model is estimated using farm household data from a sample of both organic and conventional Vietnamese farms specialized in tea production. Data was gathered through a survey questionnaire and face to face interviews in 2013 in Vietnam.

The dependent variable in the logit model is the farm household's choice to adopt organic tea cultivation. We consider several explanatory variables involving farm household and farm characteristics as well as exogenous and territorial factors. The empirical findings are consistent with the theoretical predictions. Four determinants were identified to increase the likelihood of adopting organic farming. Farm households with higher revenues, belonging to a tea association, being consulted by extension agents and located in Phu-Tho province are significantly more likely to adopt organic tea farming. The results of the basic model are largely robust and are not influenced by unobserved heterogeneity and endogeneity. Increasing the sample size by applying multiple imputation method could not alter the quality of the results.

In the last years, the Vietnamese government has set up specialized agricultural extension programs and practical training measures to encourage farm households to develop organic farming. These policy actions have proven to be very effective and play a crucial role in the support for organic tea cultivation. Our results about the potential endogeneity of the variable EXTENSION confirms this observation. On the other hand, other forms of support need to be improved, such as guaranteeing a fixed organic tea price, relieve certification costs or helping farmers to distribute and export organic tea products. Further efforts have been put forward to address the lacking marketability and distribution of organic products. The certification scheme *VietGAP* has been initialized by the Ministry of Agriculture and Rural Development in 2008 to increase the quality and enhance the exports of food products (Ha, 2014; Van Bac et al., 2017). However, third party certification schemes are cost-intensive for producers/farmers and the standard obligations are very often administratively burdensome. As a consequence, small-scale farmers and poor farm households are more reluctant to switch to organic farming and are not inclined to comply with certification standards (Van Bac et al., 2017). An alternative solution, named Participatory Guarantee System (PGS), represents the direct participation of farmers, producers, government, private sector, supporting organizations (NGOs) and consumers in the process to guarantee the quality of safe agricultural products. This form of quality assurance system builds on trust, social networks and knowledge exchange and is adapted to local markets and short supply chains (Ha, 2014; IOFAM, 2013; Willer and Lernoud, 2016). Our findings reveal that poorer farm households are less prone to adopt organic farming. Alternative policy packages need to be addressed to support those farmers who do not have sufficient financial capital to bear the risks of adopting new agricultural practices.

Several limitations of our analysis need to be highlighted. It must be acknowledged that we chiefly focus on the production side of the organic tea farming. To provide a better understanding about the potential of the domestic and foreign markets it would be beneficial to analyze market access, the distribution channels, the (local) food chain and consumer willingness to pay for organic tea products. Furthermore, due to the limitation of the questionnaire, several explanatory variables are not present in our analysis such as wage labor forces, environmental concern, and subsidies to produce organic tea. To overcome these limitations, more research appears necessary.

In a next step, future research should aim to extend the present work in order to improve the current methodologies with respect to unobserved heterogeneity, endogeneity and missing values. The full potential has not yet been sufficiently explored in this realm and it seems necessary to construct more accurate and unifying metrics of predictions by means of Bayesian analysis. It would be beneficial to compare the advantages and drawbacks of alternative metrics (e.g. a metric that considers the accuracy of predictions across separate components of the sample) with our measure.

Provisions of agricultural extension activities are crucial for subsistence farm settings having no access to other intellectual capital (i.e experience, learning by-doing and formal education). With its importance, it has spurred a debate among scholars on how to assess the value of agricultural extension services. A set of studies explores theoretically and empirically the supply and demand for distinct forms of agricultural extension systems (Dinar and Keynan, 2001, Frisvold et al., 2001, Hanson and Just, 2001 Holloway and Ehui, 2001, Kidane and Worth, 2016). The paper by Ainembabazi and Mugisha (2014) found a non-linear relationship between farmers know-how and adoption of agricultural technology, which has also become of interest in industrial economy (Aghion et al., 2005). Extending the present work to the evaluation of the value of extension service provisions could therefore be a fruitful avenue for future research.[14]

[14] We thank the anonymous reviewer for his helpful thoughts on the evaluation of the value of extension

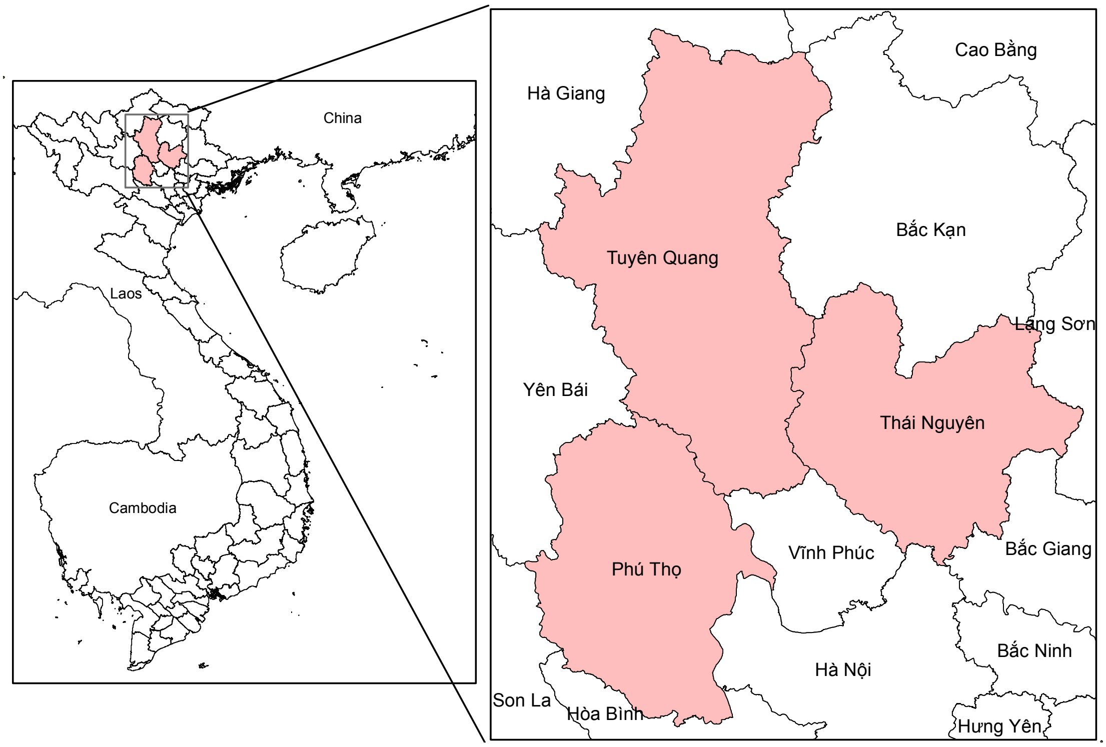


Figure 1: Geographical Location of the three Provinces in Northern Vietnam. Source: Adapted from the Vietnam Ministry of Natural Resources and Environment.

service provisions.

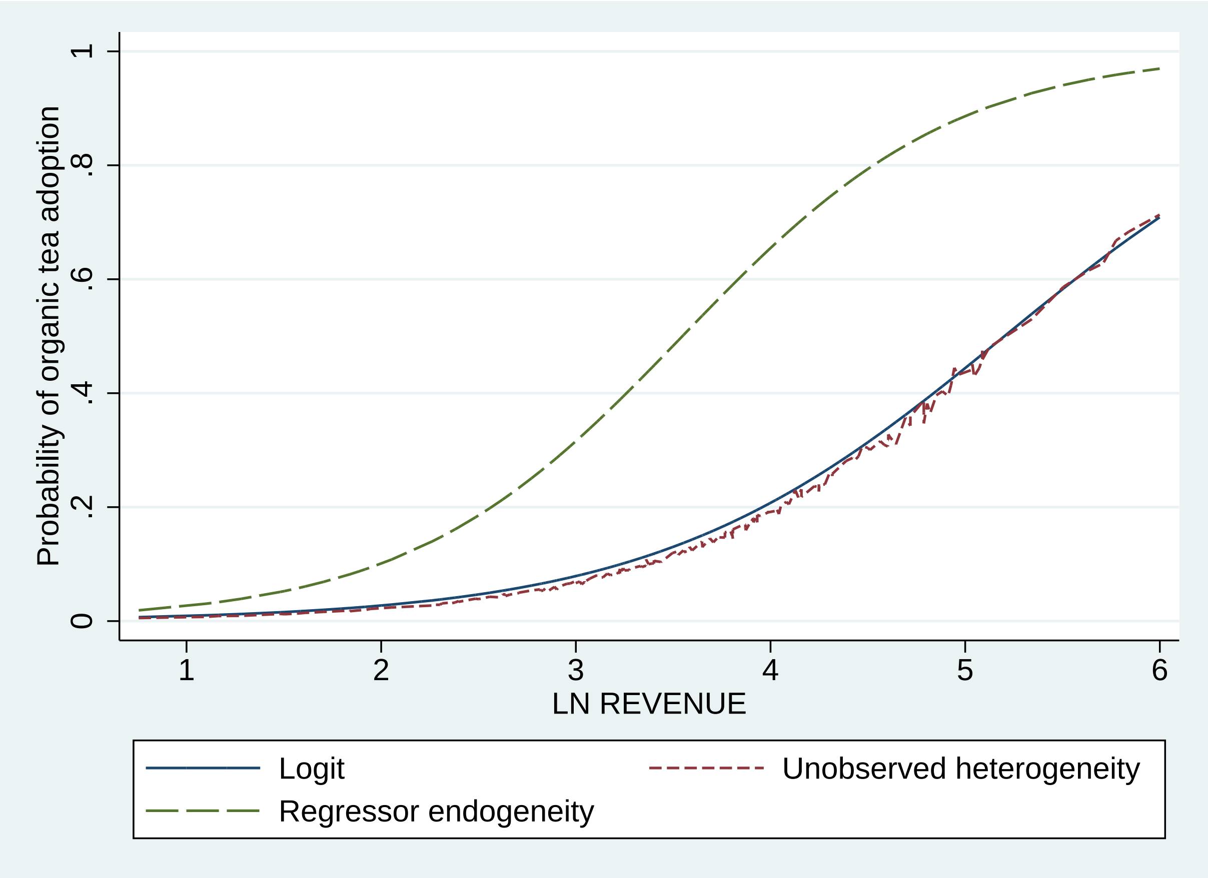
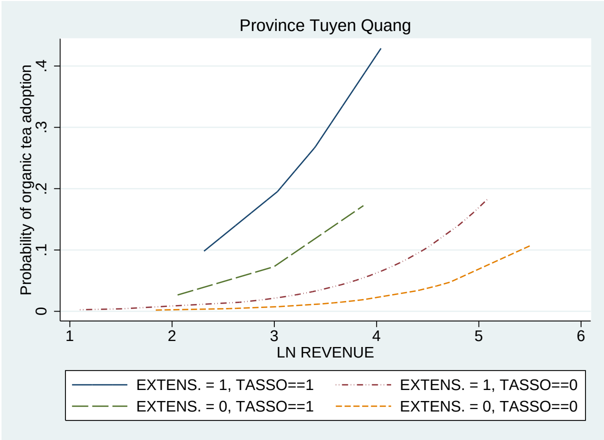
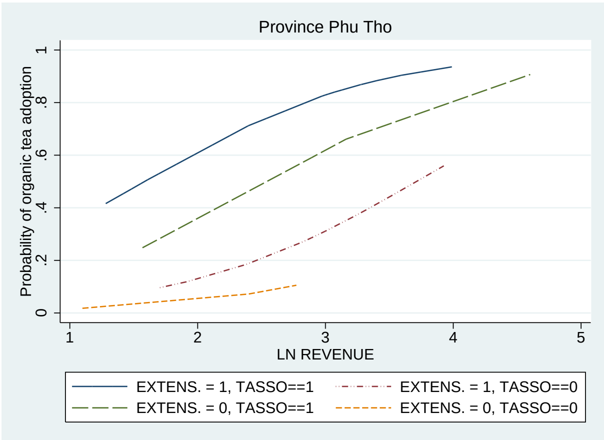


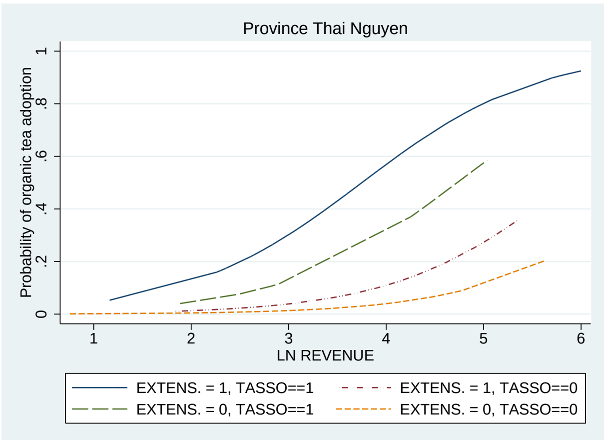
Figure 2: Relation between Probability of Organic Tea Adoption and Household Revenue (using the sample mean of other explanatory variables).



(a) Tuyen Quang Province



(b) Phu-Tho Province



(c) Thai Nguyen Province

Figure 3: Predicted Probability of Organic Tea Adoption, Extension Activity and Membership of a Tea Association using Logit Model.

Table 1: Summary of Variables included in the Estimations

Variable name	Definition	Nature	Mean	Std. Dev.	Min.	Max.	Obs.
CHOICE	Cultivation of organic tea		0.20	0.40	0	1	216
	(1 if yes, 0 otherwise)	dummy					
REVENUE	Farm household revenues per cultivated tea area (million VND)	continuous	65.78	67.09	2.13	403	241
LAND	Farm size measured in hectares	continuous	0.58	0.60	0.01	4.5	241
LABOR	Total labor employed in tea production (persons per day)	continuous	225.01	545.75	5	7863	241
EXPERIENCE	Years of experience in tea cultivation of the household head	continuous	29.75	13.84	2	64	241
HHSIZE	Household size (persons living in the household)	continuous	4.35	1.14	2	10	239
HEDUC	High educational level of the household head (high school or above)		0.33	0.47	0	1	241
	(1 if yes, 0 otherwise)	dummy					
MINORITY	Head of the household corresponds to a minority		0.11	0.31	0	1	241
	(1 if yes, 0 otherwise)	dummy					
GENDER	Head of the household's gender		0.43	0.50	0	1	241
	(1 if male, 0 otherwise)	dummy					
TASSO	Farm household is member of a tea association		0.37	0.48	0	1	215
EXTENSION	Farm household has been consulted by extension agents		0.72	0.46	0	1	241
	(1 if yes, 0 otherwise)	dummy					
$FERT_{CH}$	Farm household applies only chemical fertilizer		0.74	0.44	0	1	241
	(1 if yes, 0 otherwise)	dummy					
$FERT_{ORG}$	Farm household applies only organic fertilizer		0.48	0.50	0	1	240
	(1 if yes, 0 otherwise)	dummy					
TUYEN-QUANG	The farm is located in the Tuyen-Quang province		0.30	0.46	0	1	241
	(1 if yes, 0 otherwise)	dummy					
PHU-THO	The farm is located in the Phu-Tho province		0.13	0.34	0	1	241
	(1 if yes, 0 otherwise)	dummy					
THAI-NGUYEN	The farm is located in the Thai-Nguyen province		0.57	0.49	0	1	241
	(1 if yes, 0 otherwise)	dummy					

Table 2: Estimation Results for the Logit Model

Variable	Coefficient	Robust Std.Err.	Marginal effect	Std.Err.
LN REVENUE	1.116**	0.426	0.088**	0.031
LN LAND	0.193	0.405	0.015	0.032
MINORITY	1.271	0.916	0.100	0.072
HEDUC	-0.0884	0.497	-0.007	0.038
GENDER	-0.116	0.536	-0.009	0.043
HHSIZE	0.174	0.167	0.014	0.013
EXTENSION	1.091*	0.567	0.086*	0.047
EXPERIENCE	-0.019	0.0179	-0.001	0.001
TASSO	2.373**	0.602	0.186**	0.054
TUYEN-QUANG	-0.612	0.943	-0.048	0.073
PHU-THO	2.417**	0.914	0.190**	0.075
Intercept	-8.110**	2.751	–	–
Number of observations	203			
Ln Likelihood	-61.514			
Adjusted McFadden's pseudo R^2	0.327			
Model's significance: Wald test[a]	$\chi^2(11) = 33.26$, p-value = 0.001			
Endogenous regressors: Wald test[b]	$\chi^2(2) = 1.86$, p-value = 0.39			
Unobserved heterogeneity: LR test[c]	$\chi^2(1) = 0.089$ p-value = 1			

Notes. Significant level: ** 5%, * 10%. [a] Test for the model's significance with the null hypothesis that all the coefficients (except the intercept) are jointly zeros (under the null the test statistic follows a $\chi^2(11)$). [b] Test for endogeneity of LN REVENUE and EXTENSION (under the null of absence of endogeneity, the test statistic follows a $\chi^2(2)$). [c] Test for existence of unobserved heterogeneity (under the null of absence of unobserved heterogeneity, the test statistic follows a $\chi^2(1)$.

Table 3: Information Criteria

Data	Criteria	Logit	Unobserved heterogeneity	Regressor endogeneity
Original data	AIC	147.03	148.69	149.25
	BIC	186.79	191.77	195.57
Imputation data	AIC	220.63	221.73	215.22
	BIC	262.45	267.03	264.00

Table 4: Predictions (number of observations in parentheses).

Model	Prediction	Observation		Predictive score[a]
		Conventional tea	Organic tea	
		Original data		
Logit	0	0.9509	0.45	0.872
		(155)	(18)	
	1	0.0491	0.55	
		(8)	(22)	
Heterogeneity	0	0.9325	0.25	0.896
		(152)	(10)	
	1	0.0675	0.75	
		(11)	(30)	
Endogeneity	0	0.5864	0.075	0.653
		(95)	(3)	
	1	0.4136	0.9250	
		(67)	(37)	
		Imputation data		
Logit	0	0.9040	0.4688	0.805
		(160)	(30)	
	1	0.0960	0.5313	
		(17)	(34)	
Heterogeneity	0	0.7966	0.2031	0.797
		(141)	(13)	
	1	0.2034	0.7969	
		(36)	(51)	
Endogeneity	0	0.4237	0.0469	0.564
		(75)	(3)	
	1	0.5763	0.9531	
		(102)	(61)	

Note. Prediction = 0 (conventional tea), 1 (organic tea).

References

Abdoulaye, T., A. Abass, B. Maziya-Dixon, G. Tarawali, R. Okechukwu, J. Rusike, A. Alene, V. Manyong, and B. Ayedun (2014). Awareness and adoption of improved cassava varieties and processing technologies in Nigeria. *Journal of Development and Agricultural Economics 6*(2), 67–75.

Abebaw, D. and M. G. Haile (2013). The impact of cooperatives on agricultural technology adoption: Empirical evidence from ethiopia. *Food policy 38*, 82–91.

Abebe, G., J. Bijman, S. Pascucci, and O. Omta (2013). Adoption of improved potato varieties in Ethiopia: The role of agricultural knowledge and innovation system and smallholder farmers' quality assessment. *Agricultural Systems 122*, 22–32.

Adebayo, S. and O. Oladele (2013). Vegetable farmers' attitude towards organic agriculture practices in South Western Nigeria. *Journal of Food, Agriculture and Environment 11*(2), 548–552.

Adesina, A. (1996). Factors affecting the adoption of fertilizers by rice farmers in Cote d'Ivoire. *Nutrient Cycling in Agroecosystems 46*(1), 29–39.

Adesina, A. and M. Zinnah (1993). Technology characteristics, farmers' perceptions and adoption decisions: A tobit model application in Sierra Leone. *Agricultural Economics 9*(4), 297–311.

Adesina, A. A. and J. Baidu-Forson (1995). Farmers' perceptions and adoption of new agricultural technology: Evidence from analysis in burkina faso and guinea, west africa. *Agricultural Economics 13*(1), 1–9.

Adesina, A. A. and J. Chianu (2002). Determinants of farmers' adoption and adaptation of alley farming technology in Nigeria. *Agroforestry Systems 55*(2), 99–112.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics 120*(2), 701–728.

Ahmed, M. H. and H. M. Mesfin (2017). The impact of agricultural cooperatives membership on the wellbeing of smallholder farmers: empirical evidence from eastern ethiopia. *Agricultural and Food Economics 5*(1), 6.

Ainembabazi, J. H. and J. Mugisha (2014). The role of farming experience on the adoption of agricultural technologies: Evidence from smallholder farmers in uganda. *Journal of Development Studies 50*(5), 666–679.

Akinola, A., A. Alene, R. Adeyemo, D. Sanogo, A. Olanrewaju, C. Nwoke, and G. Nziguheba (2010). Determinants of adoption and intensity of use of balance nutrient management systems technologies in the northern Guinea savanna of Nigeria. *Quarterly Journal of International Agriculture 49*(1), 25–45.

Alem, Y. and N. H. Broussard (2018). The impact of safety nets on technology adoption: a difference-in-differences analysis. *Agricultural Economics 49*(1), 13–24.

Ali, A. and A. Abdulai (2010). The adoption of genetically modified cotton and poverty reduction in Pakistan. *Journal of Agricultural Economics 61*(1), 175–192.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature 19*(4), 1483–1536.

Arslan, A., N. McCarthy, L. Lipper, S. Asfaw, and A. Cattaneo (2014). Adoption and intensity of adoption of conservation farming practices in Zambia. *Agriculture, Ecosystems & Environment 187*, 72–86.

Ayuk, E. (1997). Adoption of agroforestry technology: The case of live hedges in the Central Plateau of Burkina Faso. *Agricultural Systems 54*(2), 189–206.

Azur, M. J., E. A. Stuart, C. Frangakis, and P. J. Leaf (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Mthods in Psychiatric Research 20*(1), 40–49.

Banerjee, B. (1981). An analysis of the effect of latitude, age and area on the number of arthropod pest species of tea. *Journal of Applied Ecology*, 339–342.

Banson, K. E., N. C. Nguyen, and O. J. Bosch (2018). A systems thinking approach to the structure, conduct and performance of the agricultural sector in ghana. *Systems Research and Behavioral Science 35*(1), 39–57.

Baumgart-Getz, A., L. Prokopy, and K. Floress (2012). Why farmers adopt best management practice in the United States: A meta-analysis of the adoption literature. *Journal of Environmental Management 96*(1), 17–25.

Birkhaeuser, D., R. E. Evenson, and G. Feder (1991). The economic impact of agricultural extension: A review. *Economic development and cultural change 39*(3), 607–650.

Bryan, E., T. T. Deressa, G. A. Gbetibouo, and C. Ringler (2009). Adaptation to climate change in ethiopia and south africa: options and constraints. *Environmental science & policy 12*(4), 413–426.

Burton, M., D. Rigby, and T. Young (2003). Modelling the adoption of organic horticultural technology in the uk using duration analysis. *Australian Journal of Agricultural and Resource Economics 47*(1), 29–54.

Caswell, M. F. and D. Zilberman (1986). The effects of well depth and land quality on the choice of irrigation technology. *American Journal of Agricultural Economics 68*(4), 798–811.

Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research 84*(4), 487–508.

Chirwa, E. (2005). Adoption of fertiliser and hybrid seeds by smallholder maize farmers in Southern Malawi. *Development Southern Africa 22*(1), 1–12.

Chuc, N., P. Singh, K. Srinivas, A. Ramakrishna, N. Chinh, N. Thang, S. Wani, and T. Long (2006). Yield gap analysis of major rainfed crops of northern vietnam using simulation modeling. global theme on agroecosystems report.

Croppenstedt, A., M. Demeke, and M. M. Meschi (2003). Technology adoption in the presence of constraints: the case of fertilizer demand in Ethiopia. *Review of Development Economics 7*(1), 58–70.

Dam, N. D. (2016). Production and Supply Chain Mangament of organic food in Vietnam. Food and Fertilizer Technology Center. URL: http://www.fftc.agnet.org retrieved on 02/10/2017.

Darnhofer, I., W. Schneeberger, and B. Freyer (2005). Converting or not converting to organic farming in Austria: Farmer types and their rationale. *Agriculture and Human Values 22*(1), 39–52.

Dey, M., F. Paraguas, P. Kambewa, and D. Pemsl (2010). The impact of integrated aquaculture–agriculture on small-scale farms in southern Malawi. *Agricultural Economics 41*(1), 67–79.

Dinar, A. and G. Keynan (2001). Economics of paid extension: Lessons from experience in nicaragua. *American Journal of agricultural economics 83*(3), 769–776.

Fairweather, J. and H. Campbell (2003). Environmental beliefs and farm practices of New Zealand farmers Contrasting pathways to sustainability. *Agriculture and Human Values 20*(3), 287–300.

FAO (2012a). Baseline survey of phu tho province. Food and Agriculture Organization.

FAO (2012b). *Statistical Yearbook 2011: World Food and Agriculture*. Food and Agriculture Organziation of the United Nations, Rome, Italy.

FAO (2015). *World Tea Production and Trade. Current and Future Development*. Food and Agriculture Organziation of the United Nations.

Feder, G., R. Just, and D. Zilberman (1985). Adoption of agricultural innovations in developing countries: A survey. *Economic Development and Cultural Change 33*(2), 255–298.

Flaten, O., G. Lien, M. Koesling, P. S. Valle, and M. Ebbesvik (2005). Comparing risk perceptions and risk management in organic and conventional dairy farming: Empirical results from Norway. *Livestock Production Science 95*(1), 11–25.

French development agency (2004). Tea growing in Phu Tho province. URL: http://www.ffem.net retrieved on 04/10/2017.

Frisvold, G. B., K. Fernicola, and M. Langworthy (2001). Market returns, infrastructure and the supply and demand for extension services. *American Journal of agricultural economics 83*(3), 758–763.

Ghosh, S. (2011). Economic implications of imputation in agricultural economic data: Discussion. *American Journal of Agricultural Economics 93*(2), 627–628.

Gómez-Carracedo, M., J. Andrade, P. López-Mahía, S. Muniategui, and D. Prada (2014). A practical comparison of single and multiple imputation methods to handle complex missing data in air quality datasets. *Chemometrics and Intelligent Laboratory Systems 134*, 23–33.

Green, G., D. Sunding, D. Zilberman, and D. Parker (1996). Explaining irrigation technology choices: a microparameter approach. *American Journal of Agricultural Economics 78*(4), 1064–1072.

GSO (2011). *Statistical Yearbook 2011.* Statistical Publishing House, Hanoi, Vietnam.

Ha, T. M. (2014). Effectiveness of the Vietnamese Good Agricultural Practice (VietGAP) on Plant Growth and Quality of Choy Sum (Brassica rapa var. parachinensis) in Northern Vietnam. *Aceh International Journal of Science and Technology 3*(3), 80–87.

Hanson, J. C. and R. E. Just (2001). The potential for transition to paid extension: Some guiding economic principles. *American Journal of agricultural economics 83*(3), 777–784.

Hattam, C. E., D. J. Lacombe, and G. J. Holloway (2012). Organic certification, export market access and the impacts of policy: Bayesian estimation of avocado smallholder ‘times-to-organic certification’ in Michoacán Mexico. *Agricultural Economics 43*(4), 441–457.

Haugen, M. and B. Brandth (1994). Gender differences in modern agriculture: The case of female farmers in Norway. *Gender and Society 8*(2), 206–229.

Heitjan, D. F. and R. J. Little (1991). Multiple imputation for the fatal accident reporting system. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 40*(1), 13–29.

Hiebert, L. (1974). Risk, learning, and the adoption of fertilizer responsive seed varieties. *American Journal of Agricultural Economics 56*(4), 764.

Holloway, G., D. J. Lacombe, and T. M. Shaughnessy (2014). How large is congressional dependence in agriculture? Bayesian inference about ‘scale’ and ‘scope’ in measuring a spatial externality. *Journal of Agricultural Economics 65*(2), 463–484.

Holloway, G. J. and S. K. Ehui (2001). Demand, supply and willingness-to-pay for extension services in an emerging-market setting. *American Journal of agricultural economics 83*(3), 764–768.

Horton, N. and K. Kleinman (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American statistician 61*(1), 79–90.

Horton, N., S. Lipsitz, and M. Parzen (2003). A potential for bias when rounding in multiple imputation. *The American Statistician 57*(4), 229–232.

Hsieh, S.-C. (2005). *Organic farming for sustainable agriculture in Asia with special reference to Taiwan experience.* Research Institute of Tropical Agriculture and International

Cooperation, National Pingtung University of Science and Technology, Pingtung, Taiwan.

IOFAM (2013). Participatory guarantee systems Vietnam. URL: https://www.ifoam.bio/en/pgs-vietnam retrieved on 12/10/2017.

Jansen, K. (2000). Labour, livelihoods and the quality of life in organic agriculture in Europe. *Biological Agriculture and Horticulture 17*(3), 247–278.

Kafle, B. and P. Shah (2012). Adoption of improved potato varieties in Nepal: A case of Bara district. *Journal of Agricultural Sciences 7*(1), 14–22.

Kalaycioglu, O., A. Copas, M. King, and R. Z. Omar (2016). A comparison of multiple-imputation methods for handling missing data in repeated measurements observational studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 179*(3), 683–706.

Kan, I., A. Kimhi, Z. Lerman, et al. (2006). Farm output, non-farm income, and commercialization in rural georgia. *eJADE: electronic Journal of Agricultural and Development Economics 3*(2).

Karki, L., R. Schleenbecker, and U. Hamm (2012). Factors influencing a conversion to organic farming in Nepalese tea farms. *Journal of Agriculture and Rural Development in the Tropics and Subtropics 112*(2), 113–123.

Kebede, Y., K. Gunjal, and G. Coffin (1990). Adoption of new technologies in Ethiopian agriculture: The case of Tegulet-Bulga district Shoa province. *Agricultural Economics 4*(1), 27–43.

Khaledi, M., S. Weseen, E. Sawyer, S. Ferguson, and R. Gray (2010). Factors influencing partial and complete adoption of organic farming practices in Saskatchewan, Canada. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie 58*(1), 37–56.

Kidane, T. T. and S. H. Worth (2016). Different agricultural extension systems implemented in africa: A review. *Journal of Human Ecology 55*(3), 183–195.

Kleinke, K. (2017). Multiple imputation under violated distributional assumptions: A systematic evaluation of the assumed robustness of predictive mean matching. *Journal of Educational and Behavioral Statistics 42*(4), 371–404.

Koesling, M., O. Flaten, and G. Lien (2008). Factors influencing the conversion to organic farming in norway. *International Journal of Agricultural Resources, Governance and Ecology 7*(1-2), 78–95.

Kuminoff, N. V., A. Wossink, et al. (2005). Valuing the option to convert from conventional to organic farming. In *AAEA Annual Meeting, Providence.*

Langyintuo, A. and C. Mungoma (2008). The effect of household wealth on the adoption of improved maize varieties in Zambia. *Food Policy 33*(6), 550–559.

Läpple, D. (2010). Adoption and abandonment of organic farming: An empirical investigation of the Irish drystock sector. *Journal of Agricultural Economics 61*(3), 697–714.

Läpple, D. and H. Kelley (2015). Spatial dependence in the adoption of organic drystock farming in Ireland. *European Review of Agricultural Economics 42*(2), 315–337.

Li, P., E. Stuart, and D. Allison (2015). Multiple imputation: A flexible tool for handling missing data. *Jama 314*(18), 1966–1967.

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association 83*(404), 1198–1202.

Ma, W. (2016). *The Impact of Agricultural Cooperatives on the Adoption of Technologies and Farm Performance of Apple Farmers in China*. Ph. D. thesis, Christian-Albrechts Universität Kiel.

Maddala, G. (1991). A perspective on the use of limited-dependent and qualitative variables models in accounting research. *The Accounting Review 66*(4), 788–807.

McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics 15*(5), 447–470.

Morris, T. P., I. R. White, and P. Royston (2014, Jun). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology 14*(1), 75.

Moss, C. and A. Mishra (2011). Imputing missing information in the estimation of production functions and systems. *American Journal of Agricultural Economics 93*(2), 619–626.

Negatu, W. and A. Parikh (1999). The impact of perception and other factors on the adoption of agricultural technology in the Moret and Jiru of Ethiopia. *Agricultural Economics 21*(2), 205–216.

Nelson, R. and E. Phelps (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review 56*(1/2), 69–75.

Nguyen-Van, P., F. Laisney, and U. Kaiser (2004). The performance of German firms in the business-related service sector. *Journal of Business and Economic Statistics 22*(3), 274–295.

Ninh, N. T. H., S. Nguyen Van, and P. Lebailly (2017). Needs of tea growers for participating in tea production insurance: a case study in phu tho province, vietnam. *AGROFOR International Journal 2*(2), 82–91.

Ojiako, I., V. M. Manyong, and A. Ikpi (2007). Determinants of rural farmers' improved soybean adoption decisions in northern Nigeria. *Journal of Food Agriculture and Environment 5*(2), 215–223.

Ouma, J. and H. De Groote (2011). Determinants of improved maize seed and fertilizer adoption in Kenya. *Journal of Development and Agricultural Economics 3*(11), 529–536.

Owusu, A., W. Nimo, D. Wilson, A. Alfred, F. Nsiah, H. Joyce, N. John, and D. Aliou (2013). Factors affecting the adoption and use of Nerica varieties among rice producing households in Ghana. *Asian Journal of Agriculture and Rural Development 3*(10), 721–735.

Padel, S. (2001). Conversion to organic farming: A typical example of the diffusion of an innovation? *Sociologia Ruralis 41*(1), 40–61.

Pandolfi, C., S. Mugnai, E. Azzarello, S. Bergamasco, E. Masi, and S. Mancuso (2009). Artificial neural networks as a tool for plant identification: a case study on vietnamese tea accessions. *Euphytica 166*(3), 411–421.

Rezvan, P., K. Lee, and J. Simpson (2015). The rise of multiple imputation: A review of the reporting and implementation of the method in medical research. *BMC medical research methodology 15*(1), 30.

Rose, R. and M. Fraser (2008). A simplified framework for using multiple imputation in social work research. *Social Work Research 32*(3), 171–178.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Saka, J., V. Okoruwa, B. Lawal, and S. Ajijola (2005). Adoption of improved rice varieties among small-holder farmers in South-Western Nigeria. *World Journal of Agricultural Sciences 1*(1), 42–49.

Schenker, N. and J. M. Taylor (1996). Partially parametric techniques for multiple imputation. *Computational Statistics & Data Analysis 22*(4), 425 – 446.

Schultz, T. (1975). The value of the ability to deal with disequilibria. *Journal of Economic Literature 13*(3), 827–846.

Scialabba, N. (2000). Opportunities and constraints of organic agriculture: A socioecological analysis. *Rome: FAO*.

Shiferaw, B. and S. Holden (1998). Resource degradation and adoption of land conservation technologies in the Ethiopian highlands: A case study in Andit Tid, North Shewa. *Agricultural Economics 18*(3), 233–247.

Staal, S., I. Baltenweck, M. Waithaka, T. DeWolff, and L. Njoroge (2002). Location and uptake: Integrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya. *Agricultural Economics 27*(3), 295–315.

Storstad, O. and H. Bjørkhaug (2003). Foundations of production and consumption of organic food in Norway: Common attitudes among farmers and consumers? *Agriculture and Human Values 20*(2), 151–163.

Strauss, J., M. Barbosa, S. Teixeira, D. Thomas, and R. Gomes Junior (1991). Role of education and extension in the adoption of technology: A study of upland rice and soybean farmers in Central-West Brazil. *Agricultural Economics 5*(4), 341–359.

Thang, T. C., B. Linh, and D. Thi (2015). How to support poor Vietnamese consumers to deal with food price volatility and food safety issues. *IDS Bulletin 46*(6), 84–89.

Tran, V.S., B. W. S. P. M. V. (2015). Perception of climate change and farmers' adaptation: A case study of poor and non-poor farmers in northern central of vietnam. *Journal of Basic & Applied Sciences 11*, 323–342.

Tran, N. (2009). Transition to organic tea production in the thai nguyen province, vietnam: Economic and environmental impacts. *EEPSEA Research Report* (2008-RR8).

Udensi, U., G. Tarawali, E. Favour, G. Asumugha, C. Ezedinma, B. Okoye, C. Okarter, P. Ilona, R. Okechukwu, and A. Dixon (2011). Adoption of selected improved cassava varieties among smallholder farmers in South-Eastern Nigeria. *Journal of Food, Agriculture and Environment 9*(1), 329–335.

Van Bac, H., T. Nanseki, and Y. Chomei (2017). Factors influencing tea farmers decision to adopt Vietnamese good agricultural practices in Northern Vietnam. *Journal of Agricultural Economics and Development 6*(2), 12–19.

Versteeg, M. N. and V. Koudokpon (1993). Participative farmer testing of four low external input technologies, to address soil fertility decline in benin. *Agricultural Systems 42*(3), 265–276.

Vink, G., L. Frank, J. Pannekoek, and S. Van Buuren (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica 68*(1), 61–90.

Welch, F. (1970). Education in production. *Journal of Political Economy 78*(1), 35–59.

Wenner, R. (2001). The deep roots of vietnamese tea: Culture, production and prospects for development. *Independent Study Project (ISP) Collection* (1159), 1–43.

Willer, H. and L. Kichler (2015). The word of organic agriculture–statistics and emerging trends 2005-2015. *IFOAM, Bonn, FiBL, Frick*.

Willer, H. and J. Lernoud (2016). *The world of organic agriculture. Statistics and emerging trends 2016.* Research Institute of Organic Agriculture FiBL and IFOAM Organics International.

Wollni, M. and M. Zeller (2007). Do farmers benefit from participating in specialty markets and cooperatives? the case of coffee marketing in costa rica1. *Agricultural Economics 37*(2-3), 243–248.

Wooldridge, J. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics 182*, 226–234.

Wossen, T., T. Abdoulaye, A. Alene, M. G. Haile, S. Feleke, A. Olanrewaju, and V. Manyong (2017). Impacts of extension access and cooperative membership on technology adoption and household welfare. *Journal of rural studies 54*, 223–233.

Yang, S. and J. K. Kim (2017). Predictive mean matching imputation in survey sampling. *arXiv preprint arXiv:1703.10256*.

Zeller, M., A. Diagne, and C. Mataya (1998). Market access by smallholder farmers in malawi: Implications for technology adoption, agricultural productivity and crop income. *Agricultural Economics 19*(1), 219–229.

Supplementary materials

Table 5: Number of farm households by tea province

District	Commune	Population [number of person]	Total Farm Households	Tea Farm Households
		Phu-Tho Province		
Thanh Ba	Van Linh	3350	930	633
Yen Lap	Ngoc Dong	3690	1065	801
		Tuyen Quang		
Yen Son	My Bang	12157	3151	2505
		Thai Nguyen		
Thai Nguyen	Phuc Triu	6231	1622	1508
Dai Tu	Hoang Nong	5568	1393	1113
	La Bang	3769	943	848
Song Cong	Ba Xuyen	4882	1220	855
	Binh Son	8460	2256	1579
	Tan Quang	5731	1456	1164

Table 6: Differences between Adopters and Non-Adopters using Kolmogorv-Smirnov Test.

Variable	Non-Adopters	Adopters	Differences	p_{value}
LN REVENUE	0.269	0.000	0.269	0.014**
LN LAND	0.305	-0.114	0.305	0.003***
MINORITY	0.354	0.000	0.352	1.000
HEDUC	0.054	0.000	0.054	1.000
GENDER	0.000	-0.019	0.019	0.867
HHSIZE	0.217	-0.018	0.079	0.079*
EXTENSION	0.173	0.000	0.173	0.256
EXPERIENCE	0.166	-0.074	0.166	0.300
TASSO	0.515	0.000	0.515	0.000***
TUYEN-QUANG	0.000	-0.329	0.329	0.001***
PHU-THO	0.221	0.000	0.221	0.068*
THAI-NGUYEN	0.108	0.000	0.108	0.068*

Notes. Significant level:*** 1%, ** 5%, * 10%.

Size of potential adopters represents 241 Vietnamese farm households.

Multiple imputation method

Multiple imputation is a statistical technique for dealing with missing values in a data set. Ad-hoc imputation methods (e.g. mean imputation, treating missing entries as a dummy variable) are based on implausible assumptions and impute the data only once to generate a complete database (Azur et al., 2011). Single imputation underestimates standard errors of estimates as this method assumes to know the unobserved value with certainty when it is actually unknown. The multiple imputation approach allows for uncertainty in the imputation by creating multiple predictions for each missing value. In this way a set of replacements for missing values are created through Bayesian arguments and specific properties. One important property to conduct multiple imputation requires that missing values are missing at random. This means that the probability of missingness depends on only on observed values and not on unobserved. Multiple imputation involves the specification of a parametric model for the missing data given the observed data, setting a prior distribution for the unknown model parameters, and simulating multiple independent draws from the conditional distribution of missing values by Bayes' theorem. For instance, an imputation model is fitted for a variable y containing missing values with parameter θ and covariates x with no missing data. Parametric imputation implies drawing θ from its posterior distribution, before drawing missing values of y from the posterior predictive distribution conditional on the draw θ^*.

Predictive mean matching is a semi-parametric imputation approach and relaxes some of the assumptions of parametric imputation. This method ensures the plausibility of imputed values especially when the normality assumption is violated (Schenker and Taylor, 1996). Technical descriptions about the application of predictive mean matching can be found in Kleinke (2017), Morris et al. (2014), Vink et al. (2014) and Yang and Kim (2017) and . Here, we depict the concept of predictive mean matching using a simple example. Plotted in Figure 4 is the relationship between a covariate x and the outcome y containing missing values. While the blueish dots (close to the regression line) represent observed values for both x and y, the reddish dots (close to the horizontal line (x)) capture observed values for x and missing values for y.

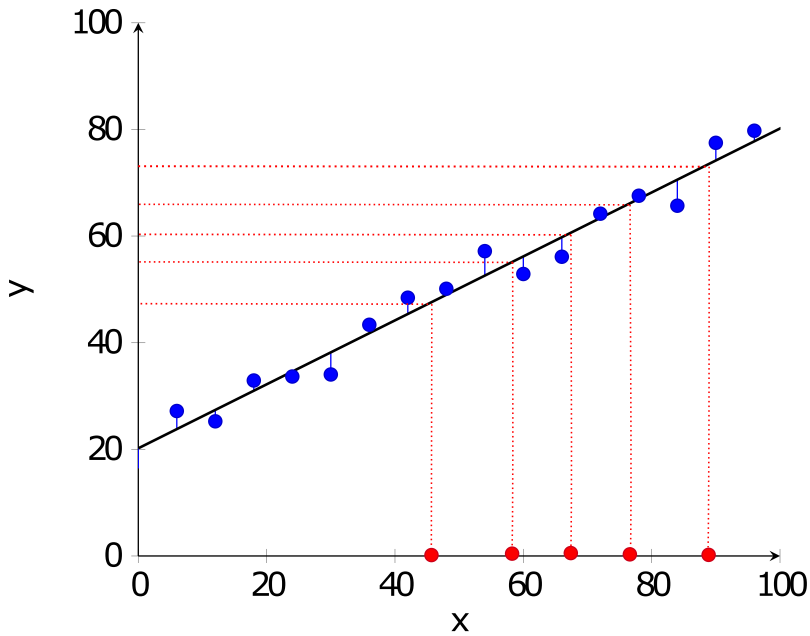


Figure 4: Concept of Predictive Mean Matching

The first step of predictive mean matching is to fit a linear regression to the observed data (only considering the observed values of both the outcome and covariate). Parametric imputation would predict the value for the missing values of the outcome based on the regression coefficients and the observed value of the covariate. By this means the missing value is replaced by the value from the predicted model. However, predictive mean matching imputes real values by using the closest observations of a donor pool matching the linear-predicted value. The donor pool is fixed[15] and contains the observed values

[15] For many statistical software the default donor pool k is equal to five. The donor pool can also be specified by using the top 5% cases in the data.

close to the predicted value. One of these values is randomly chosen to donate.

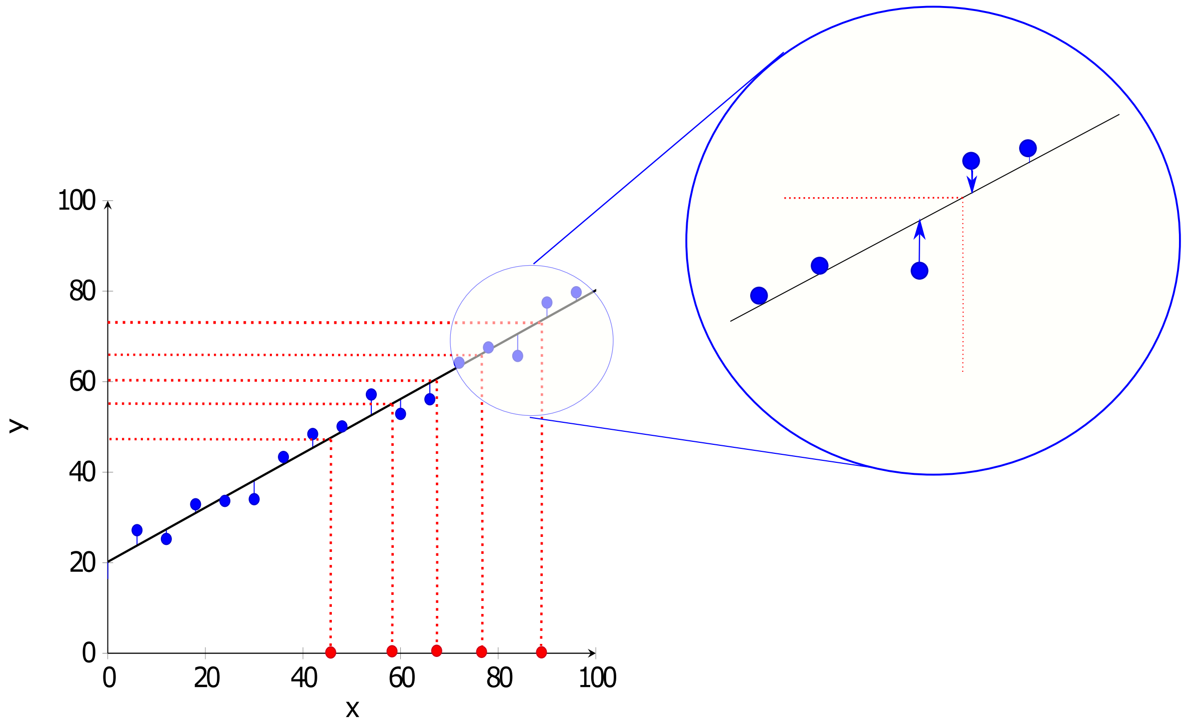
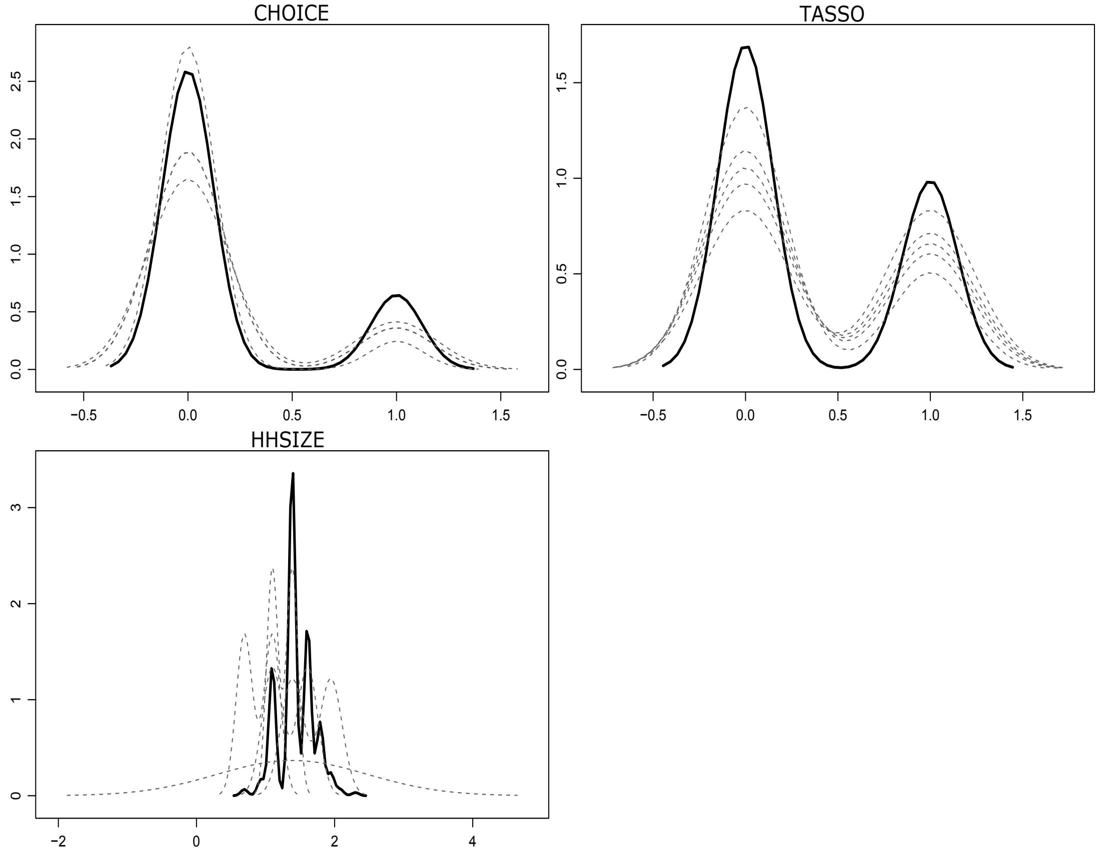


Figure 4: Concept of Predictive Mean Matching, continuous

Another important property of predictive mean matching is the overlap between observed and missing values. If there are no matches given the observed value of the covariate and the estimates from the parametric model, this method might be less appropriate. Figure 5 visualizes the density of the observed (solid line) and the imputed data (dashed line). It is important to bear in mind that the assumption of missing at random holds if the imputed and observed density distributions are akin. Figure 5 clearly demonstrates that the assumption holds for our dataset.

Figure 5: Imputation by Predictive Mean Matching



"Solid" and "dashed" line correspond to the observed and imputed data, respectively
Source: Own calculations

View publication stats