The Rural Economy Research Centre

Working Paper Series

Working Paper 10-WP-RE-05

The importance of accounting for unobserved heterogeneity, state-dependence and differences in residual variances across groups: An application to Irish Farmers land market participation decisions

STEPHEN O'NEILL[†‡] and **KEVIN HANRAHAN**[‡]
[†]Department of Economics, National University of Ireland, Galway, Ireland
[‡]Rural Economy Research Centre (RERC), Teagasc, Ireland
Stepheno_neill_1999@yahoo.com

For More Information on the RERC Working Paper Series
Email: Cathal.odonoghue@teagasc.ie,
Web: www.tnet.teagasc.ie/rerc/

The importance of accounting for unobserved heterogeneity, state-dependence and differences in residual variances across groups: An application to Irish Farmers land market participation decisions

STEPHEN O'NEILL[†‡] and KEVIN HANRAHAN[‡]
[†]Department of Economics, National University of Ireland, Galway, Ireland
[‡]Rural Economy Research Centre (RERC), Teagasc, Ireland

Abstract

Land is an essential input into agricultural production. A grwoing literature is concerned with the factors influencing farmers' land market participation decisions in developing countries, with developed countries largely ignored. Current best-practise in the land market participation literature is exemplified by Holden *et al.* (2007) who use a dynamic model which allows for state-dependence and unobserved heterogeneity. Much of the literature fails to adequately deal with these features of land market decisions. In addition, a single model is used to represent all farm types. In this paper, we firstly consider the factors influencing land market participation decisions in a developing country, Ireland, while allowing for state-dependence, unobserved heterogeneity and differences across farm tyes. We compare these results to those that are obtained while ignoring state-dependence, unobserved heterogeneity and differences between farm types. Our results suggest that some caution may be warranted when these aspects are ignored when if fact they are present.

Corresponding Author: Stephen O'Neill

Email: Stepheno_neill_1999@yahoo.com

1 Introduction

While there is an extensive literature concerned with the price that farmers are willing to pay for agricultural land, ranging from time-series analyses to the use of hedonic models, less work has been carried out on other land market decisions, such as farmers' decisions whether to participate in land rental markets or how much land to rent-in. Recent work concerned with farmers' land market participation decisions has focused on developing countries. Authors such as Skoufias (1995) have considered land market participation to be a response to imperfect markets for other inputs in the production process - such as bullocks and family labour in a developing country. This paper applies a similar model to farmers' land market decisions in a developed country, namely Ireland. However, the primary focus of the paper is on the adequacy of the econometric models commonly used in the existing literature.

In addition to being influenced by observable factors such as the use of non-land inputs, farmers' land market participation decisions may depend on a plethora of unobservable (to the econometrician at least) factors such as the farmers ability, the existence of contiguous plots of land and the farmers attitude towards lands market participation. Such factors are referred to as unobserved heterogeneity. In the existing participation literature, datasets with repeated observations of the same farm are rarely utilised, meaning that much unobserved farm heterogeneity cannot be taken account of.[i]

Thus there is an implicit assumption in much of the literature that this unobserved heterogeneity does not impact on estimates of the influence of factors affecting land market participation decisions. Wooldridge (2002) has shown that if the unobserved heterogeneity is uncorrelated with the X variables then Average Partial Effects are consistently estimated even if this heterogeity is ignored. However when the neglected heterogeneity is correlated with the explanatory variables estimated average partial effects are no longer consistent. Thus in much of the existing literature, presented estimates may not be consistent.[ii] Where farm-level panel data has been used, (Yao (2000); Holden, Deininger and Ghebru (2007); Ballesteros and Bresciani (2008)), results suggests that unobserved heterogeneity is present implying that such concerns are not mute. One aspect of our paper compares estimates of average partial effects obtained using models that incorporate unobserved heterogeneity to those obtained using models that do not.

There are also likely to be considerable costs relating to initial participation in land markets, such as search and contracting costs. Thus farmers that have already undertaken these activities in previous periods, may face lower costs of participation in the current period and so are more likely to participate again in the current period than an otherwise identical farmer that has not previously participated. This phenomenon is known as state-dependence and it may be modelled using a dynamic model. The existence of state-dependence in Irish farmers' participation decisions is strongly suggested by the fact that only approximately 5% of farms alter their decision whether to rent-in land in any given year. Although less dramatic, the area transacted in any year also displays considerable inertia.

It is not sufficient however to include a lagged dependent variable to capture this inertia as, in the presence of unobserved heterogeneity, the 'initial conditions problem' is encountered, requiring the use of a different estimator.[iii] However, with one exception that we are aware of, such methods have not been employed in the literature to date,[iv] with the result that state-dependence is routinely ignored. This paper considers the impact of ignoring state-dependence and of failing to take account of the 'initial conditions' problem. While ignoring unobserved heterogeneity and state-dependence is undesirable from a theoretical perspective, it is often necessitated by the absence of panel data across time in many developing countries. Thus it is important to ascertain the magnitude of likely biases that can result when state-dependence is ignored when modelling farmers' land market decisions.

In Ireland, the National Farm Survey, which is conducted each year by Teagasc (the Irish Agriculture and Food Development Authority), allows the construction of a panel dataset to take account of state dependence and unobserved heterogeneity in a similar manner to that of Holden, *et al.* (2007). The significance of the terms capturing state-dependence and unobserved heterogeneity in the results of Holden, *et al.* (2007) suggest that the Wooldridge estimator is to be preferred when modelling farmers' land market decisions. However, since they do not include results obtained from models that ignore state-dependence or unobserved heterogeneity, the magnitude of bias in studies which ignore these considerations is unclear. In light of the number of studies which ignore these considerations, we feel this is an unfortunate omission. In this paper, we explore the importance of incorporating unobserved heterogeneity and state-dependence in models of farmers' land market participation decisions by comparing Average Partial Effects obtained from models estimated using the Wooldridge estimator to those obtained when unobserved heterogeneity and/or state-dependence are ignored.

A further notable feature of the literature is the use of a single model to represent all farm enterprises. If farmers' in these countries are fairly homogenous in terms of their actions perhaps this is of little concern. However if farms are engaged in one of a number of possible distinct activities, it is reasonable to expect that differences may exist in the determinants of farms land market participation decisions across these activities. For example, for some farm enterprises land and labour may be complementary inputs in production, while in others they may be substitutes. To the extent that farmers engage in differing activities in developing countries and that differences in land market participation decisions exist across farm types, the estimates contained in the existing literature may prove somewhat unreliable.

In the National Farm Survey, farms are classified as belonging to one of six farm 'systems' based on the EU farm typology as set out in Commission Decision 78/463 and its subsequent amendments. Farms are categorized into the following systems: dairying, dairying and other activities, cattle-rearing, cattle and other activities, mainly sheep and tillage. Since we observe each farms system we can explore differences in the impact of influential variables across farm systems to assess whether the use of a single model for all farm types in the existing literature is likely to provide misleading estimates.

The remainder of this paper is structured as follows. In Section 2, we provide a brief theoretical model to give some context to the results, Section 3 discusses the data, and Section 4 considers the econometric methodology and discusses the statistical issues outlined above in more detail. Section 5 presents the results and Section 6 concludes.

2 Theoretical Model

According to Skoufias (1995), in a developing country, land transactions can be viewed as an attempt by farmers to adjust the area that they cultivate in response to their endowments of non-land inputs such as family labour and bullocks, which are not easily tradable due to market imperfections, in order to reach an optimal input mix.[v] Farmers may face market imperfections such as transaction costs or the absence of markets which hinder the direct adjustment of non-land inputs. An example of transaction costs typically cited in the literature (such as Skoufias (1995)) is the supervision costs associated with hired labour which make family and hired labour imperfect substitutes (under the assumption that family labour is less likely to withhold effort).

In the presence of market imperfections, farmers may alter their cultivated area to match their endowments of other inputs, rather than change their use of non-land inputs. If this were the case, households with a greater number of bullocks and availability of family labour would be expected to rent-in land so as to operate using a factor input-mix closer to their optimum. Other transaction costs include those relating to searching for tenants/land and draughting of contracts may represent barriers to participation in the land market. Skoufias assumes that, based on their initial non-land endowments of draft animals (B) and family labour (F), each farm has an optimal area they wish to farm - the desired cultivated area (DCA). Farms will participate in the land market to adjust their initial endowment of land to this level. Thus farms wish to transact an area of:

$$Y^{*} = DCA - \square = f(B,F) - Land \qquad (1)$$

However in the presence of transaction costs, farms will not transact this area but rather some function of the desired level, $Y = h(Y^{*})$. The observed land market activity, Y, is approximated using a first order Taylor-series expansion which gives:

$$Y = (h' f_B)B + (h' f_F)F - (h')\square + A \qquad (2)$$

where $h' = \frac{dh}{Y^*}$ and A is a constant. Equation 2 provides the basis for the econometric model used by Skoufias when looking at the extent of participation. We may also view this model as driving decisions whether to participate in the land market. Although in a developed country such as Ireland, one would expect that market imperfections are not as prevalent as in developing countries, in rural communities it is not hard to imagine family labour being imperfectly tradable due to barriers such as commuting costs for off-farm jobs and children below the legal working age contributing to the farm enterprise. Other authors have derived their econometric models directly from profit maximising behaviour. However the resultant econometric model takes a similar form to that implied by Skoufias model. Since we are focussed on the adequacy of the econometric models, this simpler theoretical model suffices to give some context to our treatment of farmers' decisions. We next discuss the data used in this paper.

3 Data

Each year, the Irish National Farm Survey (NFS) of approximately 1,200 Irish farmers is conducted by Teagasc (the Irish Agriculture and Food Development Authority).[vi] The survey data is nationally representative of Irish dairy, cattle and sheep farmers. In this paper we construct a panel of Irish farmers using NFS data for the years from 2000 to 2008 inclusive. In total there are 10,513 observations relating to 2,129 different farms, with approximately 1/4 of the farms being present for all nine years. To counteract the effects of influential observations, farms that reported values in the upper and lower 1% of observations for key variables for any year were excluded from the study. The final working sample consisted of 8,741 observations.

We focus on two decisions facing Irish farmers; firstly whether to participate in the land rental market and secondly, how much land, if any, to rent-in. The dependent variable for land market participation, *Renting*, is a binary variable equal to one if the farmer rents-in land in the current year and zero otherwise. When considering the extent of participation, the dependent variable, *Conacre rented*, equals the area of land rented-in under the 'conacre' system in the current year.[vii] The dependent variable for conacre rented is censored from below at zero. The same explanatory variables are used for both models and are similar to those that commonly appear in the existing literature.

Older farmers may not wish to farm as actively as younger farmers and hence may rent-in less land than younger farmers (Teklu and Lemi (2004); Holden and Ghebru (2005); Noev (2008). To capture this effect we include the age of the head of household in quadratic form, (*Age* and Age^2). Farmers that have an off-farm job may have less time to devote to agriculture and so may also rent-in less land than those without off-farm jobs (Kung (2002); Holden and Ghebru (2005)). On the other hand, income from off-farm jobs may serve to relax credit constraints so the effect is somewhat ambiguous. We include a dummy variable which equals one if the head of household[viii] has an off-farm job and zero otherwise (*Job*).

Farmers who have access to a large number of units of paid labour, (*Paid labour*) and unpaid labour, (*unpaid labour*) may be anticipated to rent-in more (less) land if land and labour are complements (substitutes) (Skoufias (1995); Deininger and Jin (2002); Tikabo, Holden and Bergland (2007)). The log of machinery operating costs is included to capture the use of machinery on the farm, (*Machinery*). The farmer's endowment of land will influence whether the farmer participates in the land market. Farmers with large initial land endowments, (*Area owned*), are expected to rent-in less land than those with small endowments (Deininger and Jin (2002); Vranken and Swinnen (2006); Ballesteros and Bresciani (2008); Jin and Deininger (2009)), though ownership of land may also relax credit constraints through the greater collateral it represents.

Since farmers with poor land quality may compensate for this by renting-in extra land, a series of dummy variables are included to capture soil quality, (*Soil1=best to Soil5=worst*). Teklu and Lemi (2004) find that land with greater erosion is more likely to be rented out, while Falkowski (2005) finds that farms with lower quality land are more likely to rent-in land, (possibly to offset its lower productivity).

Farmers' with a large number of livestock are likely to need more land for grazing etc. so a variable representing the number of livestock units is also included, (*Dairy units, Cattle units and Sheep units*). The impact of a recent policy change is captured by the inclusion of a dummy variable which equals zero in the years prior to the policy change (2000-2004 inclusive) and is equal to one from the year 2005 onwards, (*Dummy decoupling*). The effects of decoupling on farmers' land market participation decisions are the focus of O'Neill and Hanrahan (2010). Table 1 displays summary statistics for binary and scale variables respectively.

4 Econometric methodology

In much of the literature previously discussed, the decision whether to participate in the land rental market is modelled as a binary choice, with farms choosing between renting-in or not renting-in land.[ix] The extent of participation is then modelled using a censored model such as tobit, which is a standard approach in much of the literature. In this section, we briefly review the probit and tobit models before discussing the treatment of unobserved heterogeneity and state-dependence both in the existing literature and in this paper.

We can think of the decision to participate in the land market as depending on an unobserved latent variable, Y_{it}^{*}, which is a function of observed individual characteristics, X_{it} and a random component, ε_{it}. Participation occurs if the latent variable takes a value greater than zero. In the absence of unobserved individual heterogeneity and state-dependence, the latent model can be written as:

$$Y_{it}^{*}=X_{it}\beta+\varepsilon_{it}^{*} \tag{3}$$

and the observed participation decision, Y_{it}, is given by[x]:

$$Y_{it} = \begin{cases} 1; if\ Y_{it}^{*} > 0; \\ 0; if\ Y_{it}^{*} \leq 0. \end{cases} \tag{4}$$

A similar latent variable framework, albeit possibly with different coefficients, may be viewed as underlying the tobit model with Equation 4 being replaced by Equation 5 below:

$$Y_{it} = \begin{cases} Y_{it}^{*}; if\ Y_{it}^{*} > 0; \\ 0;\ \ if\ Y_{it}^{*} \leq 0. \end{cases} \tag{5}$$

In addition to their observed characteristics, X_{it}, farmers' land market participation decisions are likely to depend on some factors which are not observed by the econometrician, such as the farmers' abilities. Also, in Ireland, it is not uncommon for farmers to own multiple plots, with the result that often farmers will rent an intervening plot in order to improve the farmer's ease of access to their land. Such unobserved factors that influence farmers' land market decisions are termed unobserved heterogeneity.

4.1 Unobserved heterogeneity

In a cross-section setting, the impact of these unobserved factors will increase the variance of the error term and lead to attenuation bias in coefficient estimates. However, as discussed in Wooldridge (2002: p470-471), Average Partial Effects (APE) will be consistently estimated as long as the unobserved heterogeneity is not correlated with the included explanatory variables. Unfortunately this is unlikely to be the case in general as unobserved factors such as the farmers' abilities and their attitudes to farming are like to be correlated with the included variables, such as the farmers' endowment of land and use of other inputs. Thus estimates of APEs are likely to be inconsistent if unobserved heterogeneity is neglected.

To some extent this can be mitigated if the unobserved factors differ across regions rather than across farms[xi] through the inclusion of regional dummy variables.[xii] However, where the unobserved heterogeneity is at an individual farm level, as is likely to be the case when considering farmers' land market decisions, the inclusion of regional dummies is not sufficient to avoid inconsistent estimates. Panel data methods allow us to take account of time-invariant unobserved heterogeneity through the inclusion of individual-level effects.

We can decompose the error term, ε_{it} into two components; an unobserved individual specific effect[xiii], μ_i and a component, ν_{it} which is assumed to be identically and independently distributed (iid) over time and individuals. Thus Equation 3 becomes:

$$Y^*_{it}=X_{it}\beta+\mu_i+\nu_{it} \tag{6}$$

If we were to just ignore the unobserved heterogeneity, μ_i and pool the observations, the time-invariant unobserved individual effect would be correlated with itself over time (attenuation bias discussed previously would also be an issue). Thus explicitly considering the unobserved individual effect should lead to superior inferences than the use of cross-section or pooled models. Although many of the papers in the literature include dummy variables representing regions, allowing some unobserved heterogeneity to be captured, only those by Yao (2000), Holden, Deininger and Ghebru (2007) and Ballesteros and Bresciani (2008) control for farm-level heterogeneity. This may be partly explained by the difficulty of obtaining farm-level panel data in developing countries. Fortunately this is not an issue in Ireland. We next discuss methods of incorporating unobserved heterogeneity.

If we believe that μ_i is correlated with our explanatory variables, we should use a fixed effects estimator. However the incidental parameters problem arises whereby parameter estimates are inconsistent and suffer from a small sample bias.[xiv] Whilst this provides little challenge in a linear setting, in a non-linear setting, we must find a statistic to condition on which removes μ_i from the likelihood to overcome the incidental parameters problem.[xv] No such statistic exists for the probit and tobit models. Generally a fixed effect logit estimator is used. However a drawback of the fixed effect logit is that only observations which exhibit a change in the dependent variable for the current period contribute identifying information. Thus, in the context of this paper, only farms which have altered their participation decision during the period are useful. This is an unattractive feature when analyzing a panel dataset where the dependent variable displays a high degree of inertia, as is the case here.

If on the other hand, we do not believe that μ_i is correlated with our explanatory variables, then we may use a random effects estimator.[xvi] However, the assumption that unobserved characteristics are uncorrelated with all of the other explanatory variables seems unrealistic. Mundlak (1978) and Chamberlain (1984) relax this assumption, allowing for a correlated random effect. This involves specifying that the unobserved effect for individual *i* is partially dependent on a function of X_{it} such as the average, $\overline{X}_i$, so that $\mu_i = \overline{X}_i + \xi_{it}$. Since the Mundlak approach allows us to take account of correlation between unobserved farm heterogeneity and farm characteristics, this approach is used throughout the paper. We now discuss the treatment of unobserved heterogeneity in the literature.

Yao (2000) estimates a fixed effects logit model for farmers' decision whether to participate in land rental markets in China[xvii], finding that individual heterogeneity is present. The results from the fixed effects model differ in terms of significance from those obtained using a pooled logit model, supporting our contention that it is important to take account of individual unobserved heterogeneity when considering farmers' land market decisions.

Holden *et al.* (2007) incorporate individual effects through the use of Mundlak terms and a random effects estimator. The Mundlak terms are significant in these results suggesting firstly that unobserved heterogeneity is present and secondly that the use of a random effect would be incorrect. Ballesteros and Bresciani (2008) incorporate random effects in their probit and tobit models. However unlike Holden *et al.*, they do not appear to allow for correlation between the individual effects and other explanatory variables.

Thus the work by Holden *et al.* appears to be the only paper which adequately incorporates unobserved heterogeneity into models for both the ’participation decision’ and for the ’extent of participation’ decision. Since unobserved heterogeneity is found in all of the studies which have tried to incorporate it, some caution is probably advisable when considering the results of studies which have not done so. Holden *et al.* do not include results using random effects or ignoring unobserved heterogeneity, so the practical import of an incorrect assumption regarding the unobserved heterogeneity is unclear. To assess the importance of taking account of unobserved heterogeneity, we estimate pooled probit and tobit models and compare these estimates to those obtained using a random effects estimator excluding, and then including Mundlak terms.

We next discuss the treatment of state-dependence in this paper and in the existing literature.

4.2 State-dependence

Past participation in the land rental market may influence a farmer’s decision whether to participate in the land rental market again in the current period if participation in the previous period has altered his preferences or the prices and constraints that he faces. For example if a farmer has undertaken search or contracting costs in the previous period, these may be lower in subsequent periods. This phenomenon is known as state-dependence.

However, suppose that a farmer rented-in land in the initial period due to the influence of some unobserved heterogeneity such as the need to access a second plot of land. In subsequent periods, the farmer may again rent-in land to gain access the second plot. To the econometrician it would (incorrectly) appear that past participation had influenced the current decision, whereas in fact the current decision is caused by the unobserved existence of multiple plots). Such apparent state-dependence is thus spurious. Thus spurious state dependence is closely linked to the issue of unobserved individual heterogeneity.

From the above examples it should be clear that when examining the existence of state-dependence, care must be taken to control for the impact of individual heterogeneity. If there were no unobserved individual heterogeneity we could simply include a lagged dependent variable to capture the state-dependence. However in the presence of such heterogeneity an issue known as the "initial conditions problem" arises. The problem is caused by correlation between the unobserved heterogeneity and the lagged dependent variable.[xviii] This violates the strict exogeneity assumption with the result that state-dependence is overestimated, while the short run effects of the other explanatory variables are underestimated (Heckman, 1981a).

Approaches to deal with the initial conditions problem have been suggested by Heckman (1981a,1981b), Orme (1997, 2001) and Wooldridge (2005) in a random effects framework.[xix] Arulampalam and Stewart (2008) compare the methods of Heckman, Orme and Wooldridge, using a Monte-Carlo study for a panel of 3,000 observations (with n=500, T=6) and find that they provide similar results. Recent results from Monte Carlo experiments by Akay (2009) suggest that for panels of duration less than 5 years, Heckman's approach is superior to that of Wooldridge, while for panels from 5 to 8 years in duration the two methods tend to give similar results, with Wooldridge's approach being superior for longer panels.

State-dependence is routinely ignored in the literature. Deininger and Jin (2002) and Holden and Ghebru (2005) include a lagged dependent variable in their models in recognition of the effect that past decisions may have on current decisions. However, these results may be somewhat unreliable if unobserved heterogeneity is present due to the 'initial conditions problem'. Holden *et al.* (2007) appears to be the only work which fully incorporates state-dependence through the use of an estimator which takes account of the initial conditions problem. However, since Holden *et al.* do not present results from models that ignore state-dependence, it is not possible to assess the empirical importance of correctly accounting for state-dependence in land market participation decisions.[xx]

We next briefly outline the Wooldridge estimator which takes account of unobserved heterogeneity, state dependence and the initial conditions problem.

4.2.1 Wooldridge Estimator

Wooldridge assumes that the distribution of the unobserved component, μ, is conditional on the initial condition, Y_{i0}, and the exogenous explanatory variables. Where previously we decomposed the error term, ε into two components, μ_i and ν_{it}, Wooldridge suggests that we additionally specify μ_i as:

$$\mu_i = \alpha_0 + \alpha_1 y_{i0} + z_i \alpha_2 + a_i \tag{7}$$

Thus Wooldridge specifies the latent variable model as:

$$Y^*_{it} = z_{it}\beta + \square y_{i,t-1} + \alpha_0 + \alpha_1 y_{i0} + z_i \alpha_2 + a_i + \nu_{it} \tag{8}$$

An attractive feature of this approach is that the model above may be estimated using standard random effects probit/tobit where the regressors are $x_{it} = [1, z_{it}, y_{i,t-1}, y_{i0}, z_i]$. In practice, this estimator requires less computational time than the implementation of the Heckman approach by Stewart2006 and tends to converge more quickly also. This fact, in conjunction with the simulation results of Akay, leads us to use the Wooldridge estimator over the Heckman estimator. The Wooldridge estimator allows us to test whether the initial conditions problem arises via a simple t-test on the coefficient of the initial condition α_1. The joint significance of the Mundlak terms can be tested through the use of an F-test on the Mundlak coefficients α_2.

Table 2 summarises the models used in the literature. The last two columns show whether the authors incorporated individual effects to capture unobserved heterogeneity and whether a dynamic model is used to capture state-dependence.

4.3 Difference across farm types

Within the NFS dataset, farms are categorized as being engaged in one of six systems based on the EU farm typology as set out in Commission Decision 78/463 and its subsequent amendments. These systems are dairying, dairying and other activities, cattle-rearing, cattle-rearing and other activities, mainly sheep and tillage. The factors influencing farmer's land market participation, and their impacts, may differ across farm types. In the literature, a single model is estimated to represent all farms. If the factors influencing farmers' land market participation decisions differ across farm types or if the impact of the factors differs then the estimated coefficients will not provide a good representation of the actual impacts of the factors influencing a farmer's decision. If there are shocks which are specific to some of the farm activities, for example the outbreak of a disease such as foot and mouth, then pooling farms in different systems will also lead to autocorrelation, further compromising the reliability of estimates.

One manner in which differences between systems could be accommodated would be to interact dummy variables representing each system with the other explanatory variables. However, Ai and Norton (2003) show that, in a non-linear setting, the sign, magnitude and significance of interaction terms may all be unreliable. While Ai and Norton also present a method for taking account of this issue, in the probit model, the scale of the error term is not identified in some cases. In the tobit model information is available regarding the scale of Y_{it}^*. So this is not a concern. However such information is unavailable when we are dealing with a binary dependent variable as in the probit model (since Y_{it} depends only on the sign of Y_{it}^*). Thus in the probit model the variance of ε_{it}^* is unidentified. Rewriting the error term ε_{it}^*, as a normally distributed error term with variance unity, ε_{it}, scaled by a parameter, σ_{ε^*}, we may rewrite equation 3 as:

$$Y_{it}^* = \frac{X_{it}\beta}{\sigma_{\varepsilon^*}} + \frac{\varepsilon_{it}^*}{\sigma_{\varepsilon^*}} = X_{it}\gamma + \varepsilon_{it} \qquad (9)$$

Therefore, in the probit model, the parameter of interest is confounded with the variance of the error term. Thus the vector of coefficients which we estimate is γ rather than the parameter of interest β. In many cases this is relatively unimportant.

However, when comparing probit coefficients across groups it is necessary that the scale of the variance is the same for each group in order for the comparison to be valid (Allison (1999)).

While it is tempting to assume that only large differences in the variances matter, Monte Carlo simulations performed by Hoetker (2004) indicate that even **small** differences in residual variation can invalidate cross-group comparisons.[xxi] Williams (2009) suggests that using a heterogeneous choice model may be more appropriate, though Keele and Park (2006) suggest that if the variance equation is mis-specified, results may be even more biased than if the heteroskedasticity is ignored. Since these issues do not arise for the tobit model as σ_{ε^*} is identified, the approach taken here is to estimate the models separately for each system, avoiding the use of interaction terms and to confine comparisons between groups to the tobit results.

While in a linear model, the marginal effect of X_i on Y is given by the coefficient, β, in a non-linear models, the marginal effect of X_i , $\frac{\partial E(Y|X)}{\partial X_i}$, depends on the value of Xβ at which it is evaluated. However, when using non-linear models with panel data, the marginal effect will also depend on the value of the unobserved heterogeneity, μ_i. Wooldridge suggests using the Average Partial Effect, which consists of averaging the partial effect across the distribution of μ_i as opposed to calculating partial effects at some particular value of μ_i. The calculation of APEs and their standard errors is discussed in Mentzakis and Moro (2007).

5 Results

We now turn to the results obtained from the various model estimations. We begin by considering the results for the Wooldridge estimator, applied to the probit (participation) and tobit (extent of participation) models, contained in Tables 3 and 4 respectively. Due to the ease of interpretation and space limitations, we only present the Average Partial Effects.[xxii] The second column of each of these tables presents the results when all farm systems are pooled together - as is the case in the existing literature. The APEs for ’lagged renting’ and ’lagged area rented’ are highly statistically significant, showing that state-dependence is present. These effects are very large suggesting that ignoring state-dependence is likely to greatly overstate the short-run effects of the other explanatory variables. The initial conditions in these models are also highly significant (Initially renting/Initial area rented). Thus if the initial conditions problem were ignored and just a simple dynamic model estimated, state-dependence would be overestimated while the effects of the other explanatory variables would be underestimated. The significance of some of the Mundlak terms suggests that the use of a standard random effects estimator to treat unobserved heterogeneity would also be incorrect, as the unobserved effect would be correlated with the explanatory variables (violating the exogeneity assumption underlying random effects).

We next explored whether the use of a single set of parameters for all farm systems is warranted. The remaining columns of Tables 3 and 4 present the results when the model is ran separately for each of the six farm types. It is clear that the significance and magnitudes of the explanatory variables differ across farm types and from the model representing all farm types. Consider, for example, the effect of an off-farm job on farmers land market decisions. Using the pooled model, the average partial effects suggest that having an off-farm job has no effect on the probability of renting-in land nor on the area rented-in. However when we estimate the models separately we can see that having a job does affect the probability of renting-in land and the area rented-in for some farm systems. Thus it is clear that pooling farm systems can lead to erroneous conclusions regarding the effects of explanatory variables. To test whether these differences are statistically significant, a likelihood ratio test comparing the likelihood for the pooled model to the sum of the likelihoods from the separate models is carried out and confirms that the individual models are preferable to the pooled model. This suggests that the prevalent use of a single model to represent all farm types in the existing literature may also give cause for some concern unless there truly aren’t differences across farm types in these countries.

In light of the number of studies which have ignored the presence of unobserved heterogeneity and/or state-dependence, we next compare the results obtained using the Wooldridge estimator to those obtained from models where these features of farmers' decisions are ignored. If the estimated effects do not greatly differ from the Wooldridge results then perhaps the deficiencies in the existing literature are not too severe. We focus here on specialist dairying farmers due to space limitations.[xxiii]. Specialist dairying farmers are the largest category with 2,536 farms engaged in this activity in the final sample.

Although Table 5 presents the coefficients from the various models[xxiv], we focus instead on Table 6 which reports the Average Partial Effects for the tobit models as the basis for our comparisons.[xxv] The use of a pooled model (Column 2), ignoring state-dependence and unobserved heterogeneity, greatly overstates the average partial effects of the explanatory variables when compared to the short-run estimates obtained from the Wooldridge estimator and would also lead us to, erroneously, conclude that having an off-farm job has no effect on the area the farmer chooses to rent-in. The inclusion of random effects (Column 3) to capture the unobserved heterogeneity reduces the overstated estimates and also shows that having an off-farm job does influence land rental decisions. However, as mentioned previously, in order for the random effects results to be valid the unobserved heterogeneity must be exogenous. The inclusion of Mundlak terms (Column 4) shows that this requirement is violated here. However, although the use of Mundlak terms does alter the estimated APEs, the magnitude of these changes is relatively small suggesting that, in this case, endogeneity of the individual effects is not a great concern. The farmers' ages and the number of unpaid labour units appear to significantly influence farmers' decisions, however when state-dependence is controlled for (Column 6) we see that this is nt actually the case. Although the significance of individual Mundlak terms differs across the final three models (Columns 4-6), there is always strong evidence that the assumption of an exogenous random effect is violated.

If we depart form the Wooldridge model (Column 6) by ignoring the initial conditions problem caused by correlation between the lagged dependent variable and the unobserved heterogeneity, and estimate a simple dynamic model (Column 5) we see that the effect of the lagged dependent variable is overstated by almost a third (i.e. 0.4173/0.321). Having an off-farm job is no longer found to have a statistically significant effect at the 10% level (though it is at the 16%), while the effect of Machinery operating costs are now significant at the 10% level. Some APEs, such as those for the number of dairy units and the number of cattle units, are understated. We also see that there are changes in the significance of the average partial effects for the soil dummies and the decoupling dummy variables.

Thus we can see that the Average Partial Effects are influenced by the assumptions regarding the presence of unobserved heterogeneity and state-dependence. Models which ignore these aspects of farmers' decision making may lead to erroneous conclusions regarding the variables that are driving these decisions and the magnitudes of their influences. The routine use of the Wooldridge estimator is to be encouraged since it allows us to ascertain whether state-dependence or unobserved heterogeneity are influencing our results.

6 Conclusions

As land is an essential input into agricultural production, it is somewhat surprising that the factors influencing farmers' land market participation decisions in developed countries has been largely ignored. Land market participation decisions of farmers in developing countries have received more attention. However, possibly due to data limitations in these countries, models which make use of panel data methods have been relatively rare. Few papers have controlled for unobserved differences between farms, while serious deficiencies exist in the treatment of state-dependence in much of the literature. Current best-practise in the land market participation literature is exemplified by Holden *et al.* (2007). They account for unobserved heterogeneity and state-dependence through the use of an estimator suggested by Wooldridge (2005). However, although their results show that state-dependence and unobserved heterogeneity are important features to incorporate, they do not explore the effects of ignoring either. In light of the number of studies which ignore these features we compare results obtained using this model to those obtained from models where unobserved heterogeneity is not modelled and where state-dependence is ignored or incorrectly treated due to presence of unobserved heterogeneity. An additional feature of the literature to date has been the implicit assumption that differences between farms engaged in various enterprises do not warrant separate estimates. This assumption is tested in this paper by estimating separate models for each farm type and comparing the resulting estimates to those obtained when all farm systems are pooled in a single model.

Our results suggest that the use of a single model to represent disparate farms may lead to misleading estimates and incorrect inferences. Similarly, if the initial conditions problem is not controlled for, state dependence is overestimated (by approximately a third in this case) while other estimates also prove unreliable. Similarly, if unobserved heterogeneity is not controlled for, and correlation between the unobserved heterogeneity and the other explanatory variable accounted for, then estimates again prove unreliable. Thus results in much of the literature may prove to be unreliable if re-examined using the Wooldridge estimator as advocated by Holden *et al.* (2007) and estimating separate models where differences exist between farm enterprises as advocated here.

7 References

Ai, C. and E.C Norton. 2003. Interaction terms in logit and probit models. *Economics Letters*, 80(1):123-129.

Akay, A. 2009. The Wooldridge method for the initial values problem is simple: what about performance? *IZA Discussion Paper*, NO. 3943, Institute for the Study of Labor, Bonn.

Allison, P. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods and Research*, 28(2):186-208.

Arulampalam, W. and M.B. Stewart. 2008. Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators. The Warwick Economics Research Paper Series (TWERPS) 884, University of Warwick. Department of Economics. http://ideas.repec.org/p/wrk/warwec/884.html

Ballesteros, M.M. and F. Bresciani. 2008. Land rental market activity in agrarian reform areas: Evidence from the Philippines. *Discussion Paper Series No. 2008-06*, Philippine Institute for Development Studies.

Bartolucci, F. and Nigro, V. 2007. Approximate conditional inference for panel logit models allowing for state dependence and unobserved heterogeneity. *Working paper*. Accessed on 22nd February 2010 at http://arxiv.org/abs/math.ST/0702774

Chamberlain, G. 1984. Panel data. In ed. Z., editor, *Handbook of Econometrics*, volume 2, pages 1247-1318. Griliches and M.D. Intrilligator, Amsterdam: North Holland.

Deininger and S. Jin. 2002. Land rental markets as an alternative to government reallocation?: Equity and efficiency in the Chinese land tenure system. *Policy Research Working Paper 2930.* World Bank, Development Research Group, Rural Development, Washington, D.C.

Falkowski, J. 2005. Key determinants of land rentals in Poland. *2005 International Congress, August 23-27, 2005, Copenhagen, Denmark.* European Association of Agricultural Economists (24511)

Heckman, J.J. 1981a. Heterogeneity and state dependence. In S. Rosen, editor, *Studies in Labor Markets*, pages 91-139. University of Chicago Press.

Heckman, J.J. 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In C.F. Manski and D. McFadden, editor, *Structural Analysis of Discrete Data with Econometric Applications*, pages 179-95. MIT Press, Cambridge.

Hoetker, G. 2007. The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, 28(4):331-43.

Holden, S., Deininger, K. and H. Ghebru. 2007. Impact of land certification on land rental market participation in Tigray region, Northern Ethiopia. *Paper presented at the Nordic Conference in Development Economics in Copenhagen, June 18-19.*

http://univmail.cis.mcmaster.ca/~mentzak/papers%20for%20website/note_marg_eff.pdf

Holden, S. and H. Ghebru. 2005. Kinship, transaction costs and land rental market participation. *Working Paper, Department of Economics and Management. Norwegian University of Life Sciences.*

Jin, S. & Deininger, K. (2009) Land rental markets in the process of rural structural transformation: Productivity and equity impacts from China. *Journal of Comparative Economics*, 37, 629-646.

Keele, L. & Park, D. K. (2006) Difficult Choices: An Evaluation of Heterogeneous Choice Models. *Working Paper*, March 3, 2006.

Kung, J. K.-S. (2002) Off-farm labor markets and the emergence of land rental markets in rural China. *Journal of Comparative Economics*, 30, 395–414.

Mundlak, Y. 1978. On the pooling of time-series and cross-section data. *Econometrica,* 46(1):69-85

Noev, N. 2008. Contracts and Rental Behaviour in the Bulgarian Land Market: An Empirical Analysis. *Eastern European Economics*, 46, 43-74.

O'Neill, S. and K. Hanrahan. 2010. An exploration of the impact of decoupling of agricultural support payments on land market participation decisions of Irish farmers. *RERC, Teagasc, Working Paper.*

Orme, C.D.. 1997. The initial conditions problem and two-step estimation in discrete panel data models. *Mimeo, University of Manchester.*

Orme, C.D. 2001. Two-step inference in dynamic non-linear panel data models. *Mimeo, University of Manchester, 2001.*

Skoufias, E. 1995. Household resources, transaction costs and adjustment through land tenancy. *Land Economics.*71(1):42-56.

Stewart, M.B. redprob: A stata program for the Heckman estimation of the random effects dynamic probit model. *Mimeo,* 2006. Accessed at

http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/stewart/stata/redprob.ado

Teklu, T and A. Lemi. 2004. Factors affecting entry and intensity in informal rental land markets in southern Ethiopian highlands. *Agricultural Economics,* 30 (2): 117-128.

Tikabo, M., S. Holden and O. Bergland. 2007. Factor market imperfections and the land rental market in the highlands of Eritrea: Theory and evidence. *Revised for Land Economics.*

Vranken, L. and J. Swinnen. 2006. Land rental markets in transition: Theory and evidence from Hungary. *World Development*, 34(3):481-500.

Williams, R. 2009 Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods and Research. 37 (4): 531-559*

Wooldridge, J.M. 2002. Econometric Analysis of Cross-Section and Panel Data. Cambridge, Massachusetts: MIT Press

Wooldridge, J.M. 2005. Simple solutions to the initial conditions problem in dynamic non-linear panel-data models with unobserved heterogeneity *Journal of Applied Econometrics.* 20(1):39-54.

Yao, Y. 2000. The development of the land lease market in rural China. *Land Economics. 76(2):252-266.*

Table 1: Summary statistics for dependent and independent variables

Binary Variables	No. of obs.	Percentage of 1's	
Renting	8741	51.36	
Job	8741	28.64	
Soil class 1 (Best)	8741	33.44	
Soil class 2	8741	17.93	
Soil class 3	8741	19.11	
Soil class 4	8741	20.55	
Soil class 5 (Worst)	8741	8.97	
Dummy decoupling	8741	43.63	
Scale Variables	**No. non-zero obs.**	**Mean**	**Std. dev.**
Conacre rented	4489	42.52	53.81
Age	8741	51.74	12.16
Area owned	8695	108.54	75.14
Paid labour	3317	0.31	0.46
Unpaid labour	8721	1.19	0.48
Machinery	8554	46.74	52.84
Dairy units	3453	49.45	22.63
Cattle units	8235	40.86	27.74
Sheep units	2944	24.56	25.52

Table 2: Overview of selected papers concerned with land market participation

Year	Author	Country	Participation	Extent of Participation	Individual Effects	Static/Dynamic
1995	Skoufias	India	N/A	Tobit	No	Static
2000	Yao	China	Logit		Yes	Static
2002	Deininger and Jin	China	Probit	Tobit	No	Lagged Dependent
2002	Kung	China	N/A	Tobit	No	Static
2003	Deininger, Jin, Demeke, Adenew and Gebre-Selassie	Ethiopia	Probit	Tobit	No	Static
2004	Teklu and Lemi	Ethiopia	Probit	Heckman 2-stage	No	Static
2004	Deininger, Castagnini and Gonzalez	Colombia	Probit	Tobit	No	Static
2005	Falkowski	Poland	N/A	Tobit	No	Static
2005	Holden and Ghebru	Ethiopia	Probit	Tobit and Craggs double hurdle	No	Lagged dependent
2006	Vranken and Swinnen	Hungary	N/A	Tobit and IV tobit	No	Static
2007	Holden, Deininger and Ghebru	Ethiopia	Probit	Tobit	Yes	Lagged dependent and initial condition
2007	Deininger, Carletto and Savastano	Albania	Multinomial logit and ordered probit	N/A	No	Static
2007	Masterson	Paraguay	Logit	Tobit	No	Static
2007	Deininger and Jin	China	Ordered probit	N/A	No	Static
2007	Tikabo, Holden and Bergland	Eritrea	Multinomial logit	Heckman-Lee, Deaton, OLS and Tobit	No	Static
2008	Noev	Bulgaria	Probit	Heckman 2-stage	No	Static
2008	Nyangena	Kenya	N/.A	Tobit	No	Static
2008	Ballesteros and Bresciani	Philippines	Probit	Tobit	Yes	Static
2008	Deininger and Jin	Vietnam	Probit and ordered probit	Tobit	No	Static
2009	Deininger, Jin and Nagarajan	India	Ordered probit	N/A	No	Static

Table 3: APEs for binary participation (probit) models estimated using Wooldridge estimator by farm system

	All Systems	**Dairying**	**Dairying and other activities**	**Cattle-rearing**	**Cattle-rearing and other activities**	**Mainly Sheep**	**Tillage**
Lagged renting	0.4427***	0.5247***	0.6191***	0.6355***	0.5267***	0.3788	0.1621†
Initially renting	0.3815***	0.2982***	0.1011**	0.1613**	0.2457**	0.4127†	0.5320***
Age	0.0002	-0.0014	0.0063*	-0.0001	0.0006	0.0024	-0.0012
Off-farm job	-0.0066	-0.0513**	0.0627**	0.0155	-0.0244†	-0.0024	-0.0002
Area Owned	-0.0022***	-0.0011*	-0.0021**	-0.0051***	-0.0026***	-0.0021*	-0.0014
Unpaid labour	0.0034	0.0179	0.0663	0.0094	-0.0487	0.0483	-0.0698
Paid labour	-0.0531**	-0.0650*	0.0222	-0.203†	-0.1499*	0.1261	-0.0841†
Machinery	0.0001	0.0001	0.0002	0.0001	-0.0004	-0.0002	-0.0001
Dairy units	0.0014***	0.0028***	0.0003	----	0.6333	0.0113	-0.0059
Cattle units	0.0023***	0.0040***	0.0006	0.0032***	0.0022***	0.0008	-0.0002
Sheep units	0.0011†	0.0054†	0.0009	0.0023	0.0012	0.0008	-0.0010
Soil class 2	0.0058	0.0322†	0.0025	-0.0562**	-0.0045	-0.0033	0.0533**
Soil class 3	0.0018	0.0071	0.0253	0.0079	0.0097	-0.0197	-0.1342**
Soil class 4	0.0240**	0.0385**	0.0084	0.0220	0.0218	0.0231	-0.3902***
Soil class 5	0.0167	0.0607*	0.0463†	0.0308	-0.0210	-0.0061	----
Decoupling Dummy	0.0078†	0.0168†	-0.0176	0.0063	0.0190*	0.0018	-0.0228
Mundlak: age	-0.0012*	0.0015	-0.0073**	-0.0003	-0.0025*	-0.0044*	-0.0004
Mundlak: area owned	0.0018***	0.0002	0.0009	0.0045***	0.0020***	0.0020*	0.0011
Mundlak: unpaid labour	0.0069	-0.0276	-0.0719	-0.0315	0.0676†	-0.0657	0.0674
Mundlak: paid labour	0.0522**	0.0516	0.0177	0.0455	0.1388†	-0.4996*	0.1160*
Mundlak: machinery	0.0005***	0.0005†	0.0004	0.0005	0.0004	0.0011	0.0007†
Mundlak: dairy units	-0.0014***	-0.0027***	0.0005	-0.0017	0.0001	0.0032	0.0053
Mundlak: cattle units	-0.0017***	-0.0028***	0.0005	-0.0009	-0.0013†	-0.0001	0.0008
Mundlak: sheep units	-0.0003	-0.0020	0.0001	0.0014	-0.0005	-0.0005	0.0006

Significance level: *** = 1%, ** =5%, * =10%, † =20%,

Table 4: APEs for extent of participation decision (tobit) models estimated using Wooldridge estimator by farm system

	All Systems	**Dairying**	**Dairying and other activities**	**Cattle-rearing**	**Cattle-rearing and other activities**	**Mainly Sheep**	**Tillage**
Lagged area rented	0.2498***	0.3210***	0.3756***	0.2530***	0.2029***	0.3701***	0.1640***
Initial area rented	0.2022***	0.1883***	0.1910***	0.1342***	0.1854***	0.0869***	0.2932***
Age	-0.0029	-0.0095	0.1133	-0.0420	0.0060	0.0075	-0.0637
Off-farm job	-0.4991	-1.5753*	0.0233	0.8733†	-1.0603†	-2.5225**	2.1590
Area Owned	-0.1429***	-0.1654***	-0.2203***	-0.1397***	-0.1951***	-0.0170	-0.2220*
Unpaid labour	0.5943	-0.9028	5.5736***	0.0611	1.2349	3.6707†	-2.3343
Paid labour	-0.6171	-1.3459	2.5793	11.5483***	-6.5026**	-1.2261	-0.6522
Machinery	0.0080†	0.0138†	0.0004	-0.0218	-0.0069	-0.0636*	0.0383*
Dairy units	0.0929***	0.1599***	0.0917†	----	0.1878	-0.1773	-0.1670
Cattle units	0.1667***	0.2428***	0.0640†	0.2160***	0.1609***	0.0985	0.1752*
Sheep units	0.0400†	0.0763	0.0640	0.1278†	0.0918	-0.0247	0.1167
Soil class 2	0.5864	1.4572	-1.4924	-1.3404	-0.3657	0.4317	6.5663*
Soil class 3	1.5866**	1.2795	2.9839*	0.6348	1.7863*	0.1869	-2.5650
Soil class 4	2.1138***	2.2769**	0.0056	1.4699†	1.9455†	3.9613*	-15.0088†
Soil class 5	0.8598	3.7348**	5.2699*	0.7602	-0.0072	-1.4546	-
Decoupling Dummy	0.0429	0.9662**	-0.8754	0.0323	0.2835	0.9075	-2.9768†
Mundlak: age	-0.1151***	-0.0804	-0.1434	0.0124	-0.0967*	-0.1484†	-0.1361
Mundlak: area owned	0.1070***	0.0961***	0.1227***	0.0944***	0.1270***	0.0098	0.1856†
Mundlak: unpaid labour	0.5317	1.2933	-6.7740***	-0.5085	0.9894	-2.9936	-1.1927
Mundlak: paid labour	0.5288	1.6705	-1.2775	-21.9003***	6.4200†	-11.5269	0.3527
Mundlak: machinery	0.0333***	-0.0094	0.0780***	0.0736**	0.0177	0.0871**	0.0394
Mundlak: dairy units	-0.0730***	-0.1163***	-0.0282	-0.0298	-0.0071	0.1221	0.6133**
Mundlak: cattle units	-0.1084***	-0.1258***	0.0483	-0.1229***	-0.0480†	-0.0909	-0.1103
Mundlak: sheep units	0.0007	0.0839	-0.0178	0.0792	-0.0427	0.0269	-0.0244

Significance level: ***= 1%, **=5%, *=10%, †=20%,

Table 5: Comparison of Models for extent of participation decision (tobit) for farmers engaged in dairying system

	Pooled		Random Effects		Mundlak Effects		Dynamic		Wooldridge	
	Beta	**p-value**	**Beta**	**p-value**	**Beta**	**p-value**	**Beta**	**p-value**	**Beta**	**p-value**
Lagged area rented							0.7995	0.0000	0.6275	0.0000
Initial area rented									0.3680	0.0000
Age	1.3969	0.0030	2.3031	0.0000	2.5747	0.0000	0.3267	0.4480	0.2484	0.5700
Age squared	-0.0148	0.0020	-0.0231	0.0000	-0.0242	0.0000	-0.0035	0.4110	-0.0028	0.5170
Off-farm job	-0.5603	0.8130	-4.2647	0.0330	-4.4366	0.0270	-2.4579	0.1630	-3.1575	0.0660
Area Owned	-0.4912	0.0000	-0.4824	0.0000	-0.4918	0.0000	-0.2846	0.0000	-0.3233	0.0000
Unpaid labour	-0.8234	0.6170	3.7552	0.0530	4.3399	0.0580	-2.4424	0.3220	-1.7647	0.4620
Paid labour	-2.1258	0.3700	0.5607	0.8070	-0.5944	0.8160	-2.5408	0.3400	-2.6306	0.3100
Machinery	0.0724	0.0010	0.0202	0.2390	0.0017	0.9240	0.0339	0.0700	0.0271	0.1340
Dairy units	0.6687	0.0000	0.5681	0.0000	0.4601	0.0000	0.2706	0.0000	0.3125	0.0000
Cattle units	0.8661	0.0000	0.8427	0.0000	0.7981	0.0000	0.3835	0.0000	0.4747	0.0000
Sheep units	1.6287	0.0000	1.0765	0.0000	0.7112	0.0170	0.1532	0.6000	0.1492	0.5930
Soil class 2	-0.4732	0.8440	-0.2346	0.9620	-1.4977	0.7610	1.7317	0.4890	2.8921	0.2590
Soil class 3	13.6017	0.0000	8.7628	0.0360	10.2899	0.0140	3.9386	0.0590	2.5481	0.2540
Soil class 4	13.6984	0.0000	8.8458	0.0140	10.4094	0.0040	4.9753	0.0060	4.4493	0.0210
Soil class 5	23.4387	0.0000	12.9823	0.0270	16.9599	0.0040	8.5049	0.0060	7.1082	0.0240
Decoupling Dummy	2.3312	0.1090	3.4208	0.0000	3.4871	0.0000	1.2531	0.1510	1.8878	0.0270
Mundlak: age					-0.3489	0.0630	-0.1136	0.4010	-0.1571	0.2500
Mundlak: area owned					-0.0682	0.3000	0.0953	0.0700	0.1878	0.0000
Mundlak: unpaid labour					-2.5288	0.5680	2.2342	0.4770	2.5279	0.4140
Mundlak: paid labour					0.7817	0.8950	4.2939	0.2620	3.2651	0.4030
Mundlak: machinery					0.1093	0.0560	-0.0161	0.6230	-0.0184	0.5840
Mundlak: dairy units					0.2471	0.0410	-0.0823	0.2900	-0.2273	0.0040
Mundlak: cattle units					0.0678	0.5100	-0.0603	0.3890	-0.2459	0.0000
Mundlak: sheep units					0.8956	0.0160	0.3533	0.1490	0.1639	0.5200
Constant	-46.0412	0.0000	-66.0452	0.0000	-63.2229	0.0000	-10.4693	0.3070	-7.7407	0.4750
		standard error		**standard error**		**standard error**		**standard error**		
sigma u			31.4553	1.3570	31.2643	1.3418	11.8862	1.0418	12.5426	0.8502
sigma e	31.7334	0.6185	14.4144	0.2972	14.3653	0.2951	13.4437	0.3122	12.8780	0.2905
Rho			0.8265	0.0135	0.8257	0.0135	0.4387	0.0468	0.4868	0.0365
N	2536		2536		2536		2060		2060	
Log likelihood	-8066.6480		-7050.0095		-7038.1899		-5427.1878		-5378.0583	

Table 6: Comparison of Models for extent of participation decision (tobit) for farmers engaged in dairying system (Average Partial effects)

	Pooled		**Random Effects**		**Mundlak Effects**		**Dynamic**		**Wooldridge**	
	APE	**(p-value)**	**APE**	**(p-value)**	**APE**	**(p-value)**	**APE**	**(p-value)**	**APE**	**(p-value)**
Lagged renting							0.4173	0.0000	0.3210	0.0000
Initially renting									0.1883	0.0000
Age	-0.0122	0.6870	0.0314	0.3880	0.1119	0.0210	-0.0032	0.9570	-0.0095	0.8680
Off-farm job	-0.2542	0.8120	-1.8263	0.0290	-1.9145	0.0230	-1.2578	0.1550	-1.5753	0.0600
Area Owned	-0.2236	0.0000	-0.2121	0.0000	-0.2179	0.0000	-0.1486	0.0000	-0.1654	0.0000
Unpaid labour	-0.3749	0.6170	1.6510	0.0530	1.9232	0.0580	-1.2749	0.3220	-0.9028	0.4620
Paid labour	-0.9679	0.3690	0.2465	0.8070	-0.2634	0.8160	-1.3263	0.3400	-1.3459	0.3100
Machinery	0.0330	0.0010	0.0089	0.2390	0.0008	0.9240	0.0177	0.0700	0.0138	0.1340
Dairy units	0.3044	0.0000	0.2498	0.0000	0.2039	0.0000	0.1412	0.0000	0.1599	0.0000
Cattle units	0.3943	0.0000	0.3705	0.0000	0.3537	0.0000	0.2002	0.0000	0.2428	0.0000
Sheep units	0.7415	0.0000	0.4733	0.0000	0.3151	0.0170	0.0800	0.6000	0.0763	0.5930
Soil class 2	-0.1923	0.8430	-0.0966	0.9620	-0.6096	0.7600	0.8741	0.4940	1.4572	0.2680
Soil class 3	6.1991	0.0000	3.8673	0.0420	4.5783	0.0190	2.0335	0.0630	1.2795	0.2600
Soil class 4	6.2480	0.0000	3.9065	0.0160	4.6356	0.0050	2.5961	0.0070	2.2769	0.0230
Soil class 5	11.5452	0.0000	5.9181	0.0410	7.9291	0.0100	4.5988	0.0090	3.7348	0.0310
Decoupling Dummy	1.0648	0.1100	1.5108	0.0000	1.5522	0.0000	0.6543	0.1510	0.9662	0.0270
Mundlak: age					-0.1546	0.0630	-0.0593	0.4010	-0.0804	0.2490
Mundlak: area owned					-0.0302	0.3000	0.0497	0.0700	0.0961	0.0000
Mundlak: unpaid labour					-1.1206	0.5680	1.1662	0.4770	1.2933	0.4140
Mundlak: paid labour					0.3464	0.8950	2.2414	0.2610	1.6705	0.4020
Mundlak: machinery					0.0485	0.0560	-0.0084	0.6230	-0.0094	0.5840
Mundlak: dairy units					0.1095	0.0430	-0.0429	0.2900	-0.1163	0.0030
Mundlak: cattle units					0.0300	0.5090	-0.0315	0.3900	-0.1258	0.0000
Mundlak: sheep units					0.3969	0.0160	0.1844	0.1490	0.0839	0.5200

Table 7: Comparison of Models for binary participation decision (probit) for farmers engaged in dairying system

	Pooled		Random Effects		Mundlak Effects		Dynamic		Wooldridge	
	Beta	**p-value**	**Beta**	**p-value**	**Beta**	**p-value**	**Beta**	**p-value**	**Beta**	**p-value**
Lagged renting	-	-	-	-	-	-	3.1449	0.0000	2.3371	0.0000
Initially renting	-	-	-	-	-	-	-	-	1.6808	0.0000
Age	0.0239	0.1810	0.3284	0.0010	0.3633	0.0000	-0.0103	0.7650	0.0040	0.9380
Age squared	-0.0003	0.1310	-0.0032	0.0010	-0.0034	0.0000	-0.0001	0.7860	-0.0002	0.6810
Off-farm job	-0.1815	0.0460	-0.6983	0.0260	-0.8259	0.0090	-0.3304	0.0350	-0.5763	0.0160
Area Owned	-0.0138	0.0000	-0.0520	0.0000	-0.0360	0.0000	-0.0068	0.2410	-0.0130	0.0910
Unpaid labour	0.0021	0.9740	0.4574	0.1370	0.4344	0.2160	0.2018	0.4940	0.2103	0.5680
Paid labour	-0.0814	0.3660	-0.1078	0.7710	-0.5449	0.2000	-0.5717	0.0980	-0.7646	0.0770
Machinery	0.0024	0.0060	0.0007	0.7940	-0.0017	0.5590	0.0014	0.5740	0.0014	0.6340
Dairy units	0.0153	0.0000	0.0596	0.0000	0.0423	0.0000	0.0234	0.0040	0.0330	0.0020
Cattle units	0.0238	0.0000	0.0818	0.0000	0.0746	0.0000	0.0314	0.0000	0.0471	0.0000
Sheep units	0.0578	0.0000	0.2308	0.0000	0.3115	0.0000	0.0410	0.2350	0.0639	0.1970
Soil class 2	0.0632	0.5100	0.5834	0.4360	0.7131	0.3740	0.2107	0.2250	0.3857	0.1810
Soil class 3	0.3093	0.0000	1.1022	0.1230	1.4690	0.0540	0.1117	0.4350	0.0833	0.7280
Soil class 4	0.2513	0.0000	0.8224	0.1560	1.0549	0.0600	0.2381	0.0530	0.4617	0.0310
Soil class 5	0.5163	0.0000	1.4365	0.2780	2.2812	0.0110	0.3864	0.0700	0.7171	0.0530
Decoupling Dummy	0.0644	0.2560	0.1401	0.3220	0.1368	0.3480	0.1359	0.1840	0.1969	0.1330
Mundlak: age	-	-	-	-	-0.0574	0.0420	0.0190	0.1360	0.0174	0.3060
Mundlak: area owned	-	-	-	-	-0.0422	0.0000	0.0000	1.0000	0.0023	0.7730
Mundlak: unpaid labour	-	-	-	-	0.3151	0.6810	-0.3174	0.3260	-0.3241	0.4480
Mundlak: paid labour	-	-	-	-	0.9685	0.2710	0.3431	0.3720	0.6070	0.2460
Mundlak: machinery	-	-	-	-	0.0218	0.0190	0.0031	0.3310	0.0059	0.1800
Mundlak: dairy units	-	-	-	-	0.0326	0.1050	-0.0203	0.0110	-0.0320	0.0040
Mundlak: cattle units	-	-	-	-	0.0332	0.0830	-0.0216	0.0070	-0.0334	0.0020
Mundlak: sheep units	-	-	-	-	-0.0338	0.6200	-0.0097	0.7290	-0.0235	0.5590
Constant	-0.6875	0.1200	-8.1638	0.0010	-7.3861	0.0030	-1.6468	0.0410	-2.2426	0.0770
				Standard error		**Standard error**		**Standard error**		**Standard error**
sigma u			4.2971	0.3755	4.7163	0.3871	0.0008	0.0129	0.8778	0.1978
sigma e										
Rho			0.9486	0.0085	0.9570	0.0068	0.0000	0.0000	0.4352	0.1108
N	2536		2536		2536		2060		2060	
Log likelihood	-1484.7379		-763.5425		-751.5912		-412.4941		-399.3031	

Table 8: Comparison of Models for binary participation decision (probit) for farmers engaged in dairying system (Average Partial effects)

	Pooled		**Random Effects**		**Mundlak Effects**		**Dynamic**		**Wooldridge**	
	APE	**p-value**	**APE**	**p-value**	**APE**	**p-value**	**APE**	**p-value**	**APE**	**p-value**
Lagged renting							0.8491	0.0000	0.5247	0.0000
Initially renting									0.2982	0.0020
Age	-0.0009	0.3280	0.0029	0.2130	0.0044	0.0350	-0.0020	0.1090	-0.0014	0.2800
Off-farm job	-0.0616	0.0480	-0.1107	0.0410	-0.1122	0.0130	-0.0361	0.0440	-0.0513	0.0220
Area Owned	-0.0046	0.0000	-0.0075	0.0000	-0.0046	0.0000	-0.0007	0.2420	-0.0011	0.0860
Unpaid labour	0.0007	0.9740	0.0658	0.1430	0.0555	0.2160	0.0213	0.4940	0.0179	0.5690
Paid labour	-0.0273	0.3650	-0.0155	0.7700	-0.0696	0.2000	-0.0605	0.0990	-0.0650	0.0760
Machinery	0.0008	0.0060	0.0001	0.7950	-0.0002	0.5600	0.0001	0.5740	0.0001	0.6340
Dairy units	0.0051	0.0000	0.0086	0.0000	0.0054	0.0000	0.0025	0.0040	0.0028	0.0020
Cattle units	0.0080	0.0000	0.0118	0.0000	0.0095	0.0000	0.0033	0.0000	0.0040	0.0000
Sheep units	0.0194	0.0000	0.0332	0.0000	0.0398	0.0000	0.0043	0.2360	0.0054	0.1910
Soil class 2	0.0215	0.5090	0.0920	0.4180	0.0989	0.3500	0.0223	0.2210	0.0322	0.1750
Soil class 3	0.1030	0.0000	0.1573	0.1070	0.1832	0.0360	0.0119	0.4340	0.0071	0.7280
Soil class 4	0.0842	0.0000	0.1241	0.1500	0.1399	0.0570	0.0252	0.0520	0.0385	0.0270
Soil class 5	0.1670	0.0000	0.1910	0.1620	0.2500	0.0020	0.0408	0.0720	0.0607	0.0590
Decoupling Dummy	0.0216	0.2560	0.0201	0.3240	0.0174	0.3460	0.0144	0.1850	0.0168	0.1310
Mundlak: age					-0.0073	0.0380	0.0020	0.1370	0.0015	0.3060
Mundlak: area owned					-0.0054	0.0000	0.0000	1.0000	0.0002	0.7720
Mundlak: unpaid labour					0.0402	0.6780	-0.0336	0.3260	-0.0276	0.4510
Mundlak: paid labour					0.1237	0.2770	0.0363	0.3720	0.0516	0.2430
Mundlak: machinery					0.0028	0.0210	0.0003	0.3310	0.0005	0.1790
Mundlak: dairy units					0.0042	0.1020	-0.0021	0.0120	-0.0027	0.0040
Mundlak: cattle units					0.0042	0.0990	-0.0023	0.0070	-0.0028	0.0010
Mundlak: sheep units					-0.0043	0.6190	-0.0010	0.7290	-0.0020	0.5580

[i] Some heterogeneity may be controlled for by allowing for regional heterogeneity. However where differences of the unobserved factors exist at the farm level this is insufficient.
[ii] In fact most of the papers in the literature report coefficients rather than average partial effects, meaning that their estimates also suffer from attentuation bias.
[iii] The initial conditions problem is discussed later in this paper.
[iv] Holden, Deininger and Ghebru (2007) being the exception.
[v] For example, if the farm has a bullock but does not have enough land to maximize the return from it, there may be an incentive to rent-out the bullock, however if such markets are imperfect it may be preferable to rent-in land instead.
[vi] The NFS is part of the Farm Accountancy Data Network (FADN) of the European Union (EU).
[vii] The vast majority of renting in Ireland is carried out under the 'conacre' system
[viii] Unfortunately data for spouses working off-farm jobs is not available for some years, although it appears to be insignificant in the years for which it is present - so the spouses job is not included here.
[ix] It is possible to also include renting-out land as an alternative. However in the literature the most common approach is to estimate separate models for renting-in and renting-out. Since we are primarily concerned with the effects of neglecting unobserved heterogeneity and state-dependence, we focus on the binary choice: rent-in/don't rent-in.
[x] Where Y_{it}=1 represents a choice to participate in the market by renting in land.
[xi] Tax regimes etc. may differ across regions while being constant within them.
[xii] Many of the considered studies do include regional dummies.
[xiii] This parameter captures unobserved heterogeneity.
[xiv] Heckman 1981b discusses the incidental parameters problem.
[xv] A transformation such as differencing or demeaning is sufficient in a linear setting.
[xvi] We find the likelihood for an individual conditional on the unobserved individual effect and then integrate over the range of unobserved effects to get the individual likelihood function.
[xvii] Yao does not estimate models for the extent of particaption.
[xviii] Since the time-invariant heterogeneity in this period also influenced the previous participation decision.
[xix] Bartolucci and Nigro (2007) have also shown that a dynamic fixed effect logit model can be approximated by a quadratic exponential model.
[xx] Holden and Ghebru (2005) appears to use data from a separate sampling period and is not directly comparable
[xxi] Hoetker (2007) suggests two approaches to making comparisons across groups. Firstly, we can consider differences in the sign and significance of variables. Alternatively, we may consider differences in the ratio coefficients, so that the unknown confounding term cancels out $\frac{\alpha_r}{\alpha_s}=\frac{\beta_r}{\sigma_\varepsilon}/\frac{\beta_s}{\sigma_\varepsilon}=\frac{\beta_r}{\beta_s}$.
[xxii] The coefficients are also available from the authors.
[xxiii] Similar results for the other systems can be obtained form the authors - though in some cases convergence is not acheived, in part this may be due to the smaller sample sizes involved.
[xxiv] To make valid comparisons, coefficients should be scaled by $\sqrt{1-\square}$) to correct for difference in the variance of μ
[xxv] Tables 7 presents the coefficients for the probit model, while Table 8 presents the APEs. However, since the scale of the variance may differ across systems and is unidentified, some caution is warranted when considering the probit results. Thus we focus instead on the tobit results.