Choices, Constraints and Welfare in Rural Peru

Thesis submitted to Trinity College, Dublin in fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.D.)

Supervisor: Professor Alan Mathews, Economics Department

2006

Eoghan Garvey

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at Trinity College Dublin or at any other university. It is entirely my own work. I also declare that the library at Trinity College Dublin has permission to lend or copy the thesis upon request.

Eoghan Garvey

Summary

This thesis is a study of farm household choices, constraints and welfare in rural Peru, for 1994. The study incorporates literature reviews of farm household models and the main results concerning off-farm participation for developed and developing countries. There is also a review of the Collective utility model, as well as much shorter reviews of the literatures on separability in labour allocation and on Data Envelopment Analysis. The data used is from the 1994 Living Standards Measurement survey, conducted by the World Bank. Four sets of econometric or programming models are estimated, using a sample of about 1000 farm households. The 10 main substantive results are:

(i) the Collective model is probably appropriate generally for rural Peru for both 2 and 3 adult families; there are suggestions, though that the intra-household Pareto efficiency it implies is more likely to exist for non-Spanish than Spanish speaking families;

(ii) a farm size distribution with a mode of around 2 hectares is the most technically efficient,

(iii) the two main ways of getting out of poverty appear to be to have farm assets or to increase education – the former seems more important for the extremely poor, the latter for the slightly better off;

(iv) on-farm crop diversification is strongly negatively correlated with off-farm work participation, suggesting that these are alternative risk management strategies; crop diversification is positively associated with improved technical efficiency and negatively with allocative efficiency; crop diversification decisions are mainly handled by the males in the household;

(v) market embeddedness is negatively correlated with off-farm work participation, and positively and strongly correlated with all forms of efficiency and income; cash flow in the household, whether it comes from being embedded in the market or from off-farm work, is handled by all family members;

(vi) higher education is associated with greater off-farm participation for the better off in non-agricultural wage labour (or in self-employment, in the case of female spouses) and with higher off-farm earnings; higher education reduces the probability of being rationed in off-farm work; it is associated with greater on-farm efficiency (but only for those who are very highly educated);

(vii) being non-Spanish speaking increases the chances of working on one's own farm unpaid; family interaction is likely to conform to the Collective model; one is much more likely to be in the lowest permanent income quintile, and to be rationed in off-farm work hours;

(viii) women are less likely to work off-farm than men, and less likely to work for wages than to be self-employed; they appear to have a higher shadow value of home time irrespective of the number of children, and households with relatively more females are more likely to hire in; there is little noticeable difference in on and off-farm returns for female labour compared to male labour;

(ix) a variety of other factors determine off-farm participation, including age, on-farm experience, local demand for non– and off-farm labour, and the actual main crops; the off-farm share of total earned income declines with farm size and with permanent income (as measured by per capita expenditure); having off-farm work does appear to be associated with allocative efficiency;

(x) between approximately 25% and 45% of those who work off-farm for wages are constrained in the amount of work they can get at the going wage. Such constraints as exist seem as if they can be overcome by having assets or education.

There are also four methodological innovations in the study. These are: (i) a test for the three-decision-maker version of the Collective model; (ii) a test to distinguish rationing from transaction costs as determinants of non-separability, (iii) the use of individual data in participation and separability estimations, (iv) the use of Censored Least Absolute Deviation (CLAD) models to test for the link between allocative efficiency and off-farm hours or earnings.

Acknowledgements

I would like to thank my supervisor Professor Alan Mathews for his clear suggestions at each stage of the work on this thesis, all of which I think were acted on.

I would also like to thank those who attended the post-graduate seminar series in TCD, especially Chris Minns and Marian Risov who also provided suggestions that were acted on.

Several of my colleagues at work in Galway are owed thanks. Through Professor Michael Cuddy I was introduced to the area of agricultural economics. The Head of Department, Brendan Kennelly, has been very encouraging since he took that position. Cathal O'Donoghue organized seminars in Galway, where I also received useful comments.

Within the economics department in Galway I would also like to thank Aoife Brick who was really invaluable at the proofing and the final printing stages, as well as Niall McInerney, who helped with proofing and with ongoing discussion of some of the ideas in the thesis.

I would like to thank my daughters, Caoife, for helping with the boring parts of the bibliography, and Lia, for taking the cat every morning and allowing me a little extra sleep. I would like to thank my wife, Sande, who, as usual, has steered me in good directions. I cannot thank her enough.

Finally, I would like to thank my mother for her encouragement and love over the years, and to dedicate the thesis both to her and to the memory of my father.

Introduction
Introduction 1
Background 1
Research objective 2
Research hypotheses 3
Justification for the research 4
Theoretical framework and data 5
Limitations 6
Overview (by chapter) 7
Summary 8

Chapter 1: The farm household model
Introduction 11
1.1 The unitary farm household model – introduction 11
1.1.1 The farm household model – specification and estimation 14
1.1.2 Unitary household model – farm household as a unit 15
1.1.3 Unitary farm household model with 2 (or more) members 17
1.1.4 Home production and unpaid farm work 19
1.1.5 Modelling the determinants of off-farm participation and income in developing countries 21
1.1.6 Main findings for developing countries 26
1.1.7 Research findings in developed countries using the farm household model 29
1.1.8 Main differences between developed and developing country results 31
1.1.9 Policy and research agendas 32
1.2 The collective model: introduction 33
1.2.1 Collective household model 34
1.2.2 The collective model and the farm household labour supply literature 41
1.3 Conclusion 42

Chapter 2: Background and Data
2.1 The dataset and rural Peru 43
2.1.1 A brief description of agriculture in Peru 44
2.1.2 The LSMS data 48
2.2 Visually exploring some relationships 53
2.3 A typology of off-farm work 59
2.4 Conclusion 61

Chapter 3: Collective versus unitary model
Introduction 62
3.1 Testing for the collective model 62
3.2 Model specifications 64
3.3 Results 70
3.4 Conclusion 73

Chapter 4: Reduced form participation models
Introduction 75
4.1 Reduced form models of participation 76
4.2 Participation estimation results 79
4.2.1 Model a) Household participation in off-farm work: 80
4.2.2 Model b) The main couple's participation decisions 82
4.2.3 Model c) Individuals within the household 84
4.2.3.1 Model c) Dichotomous dependent variables 85
4.2.3.2 Model c) Trichotomous dependent variables 89
4.2.3.3 Model c) Polychotomous dependent variables 93
4.2.3.4 Model c) Two extensions 96
4.3 Discussion of main points of participation results 98
4.4 Overall discussion 102

Chapter 5: Separability, the household and off-farm labour

Introduction 105
5.1 Separability and non-separability in labour allocation decisions 107
5.1.1 Testing for non-separability 108
5.1.2 Market imperfections and on-farm production 110
5.2 Separability in the labour market – the model 112
5.3 Reduced form separability tests 114
5.4 A new test for diagnosing the reasons for non-separability 118
5.5 Heterogeneous individuals in reduced form separability testing 120
5.5.1 The econometric model 121
5.6 Conclusions 127

Chapter 6: Efficiency, permanent Income and off-farm work

Introduction 129
6.1 Off-farm work and efficiency 129
6.1.1 Data envelopment analysis and stochastic frontiers 130
6.1.2 DEA models 132
6.2 Data and variables 133
6.3 Efficiency results 134
6.4 Econometric estimations 140
6.5 Off-farm work and permanent income: introduction 145
6.6 Permanent income regressions 145
6.7 Conclusion 148

Chapter 7: Drawing the Conclusions Together

Introduction 151
7.1 Substantive Conclusions 151
7.1.1 The black box of the household 151
7.1.2 Evidence for the most efficient farm size 153
7.1.3 Effects of off-farm work on permanent income 154
7.1.4 Crop diversification 155
7.1.5 Market embeddedness (proportion of crops sold) 156
7.1.6 Education 157
7.1.7 Ethnic differences (as indicated by main language of household) 158
7.1.8 Gender 159
7.1.9 Determinants of off-farm and non-farm work, and earnings shares 160
7.1.10 Constraints on off-farm work and on hiring-in 161
7.2 Methodological innovations 161
7.2.1 The use of individual level data in participation and hours regressions (ch. 4) and in separability regressions (ch. 5) 162
7.2.2 Testing the collective model for three person families (and for families of different cultural background) using the Slutsky Matrix approach (ch. 3) 163
7.2.3 A New test to distinguish rationing from non-linear transaction costs as determinants of non-separablity (ch. 5) 163
7.2.4 Examining the relationship between allocative efficiency and off-farm work (ch. 6) 164
7.3 Possibilities for further research and implications for policy 164
7.4 Conclusion 170
Bibliography 171
Appendices 189

List of Tables

Table 1-1: Percentage of income of the rurally employed not from agriculture 12
Table 1-2: Comparative statics in the unitary utility model 16
Table 2-1: GDP and inflation growth in 1990s 45
Table 2-2: Urban / rural population shares over time 45
Table 2-3: Sectors people work in off the farm 47
Table 2-4: Farm Size, Tenure and Main Crops 48
Table 2-5: Crops grown in the three main Peruvian regions 48
Table 2-6: Main summary statistics 49
Table 2-7: Mean hourly returns for family members in four different off-farm categories 49
Table 2-8: Summary statistics by expenditure per capita quintiles 50
Table 2-9: Summary statistics by region 51
Table 2-10: Hourly wages and self-employment returns in non-agricultural sectors 52
Table 2-11: Exploratory regression on total expenditure (1) 57
Table 2-12: Exploratory regression on total expenditure (2) 59
Table 3-1: 3 sets of Hicks elasticities 73
Table 4-1: Household Participation Model 80
Table 4-2: Participation of Operator and Spouse 82
Table 4-3: Participation of Individuals 85
Table 4-4: Random effects individual participation models 87
Table 4-5: Multinomial logit individual participation (3 categories) 89
Table 4-6: Random effects multinomial logit (3 categories) 91
Table 4-7: Multinomial logit individual participation (6 categories) 93
Table 5-1: Household and individual work regimes 115
Table 5-2: Reduced form separability tests 117
Table 5-3: Testing exogeneity: instrumental variable models and chow tests 118
Table 5-5: Heterogenous individuals in constrained and unconstrained classes 124
Table 5-6: Summary statistics for the constrained and the unconstrained 125
Table 6-1: Technical, Allocative and Scale Efficiencies in the Three Regions 134
Table 6-2: CLAD efficiency regressions 141
Table 6-3: CLAD efficiency regressions without the market embeddedness variable 143
Table 6-4: Regressions for scale efficiency 144
Table 6-5: Determinants of permament income 146

List of Figures

Figure 2-1: Urban/Rural Population Shares, 1993 (source: INEI internet site) 45
Figure 2-2: Farm Size in the 3 Regions (normal distribution super-imposed) 53
Figure 2-3: Farm size and per capita expenditure 55
Figure 2-4: Education and Farm Size 55
Figure 2-5: Education and Off-Farm Hours 55
Figure 2-6: Education and per-capita expenditure 56
Figure 2-7: The number of income sources and expenditure 56
Figure 2-8: Crop diversification and per capita expenditure (LOESS line drawn in) 57
Figure 2-9: Proportion of crops sold and per capita expenditure (LOESS line drawn in) 57
Figure 2-10: Two Dimensional Correspondence between Clusters and Sectors 60
Figure 6-1: Proportion of crops sold and efficiency (with LOESS lines drawn in). 136
Figure 6-2: Risk index and efficiency (with LOESS lines drawn in) 137
Figure 6-3: Language Group and efficiency 138
Figure 6-4: Education and efficiency (with LOESS lines drawn in) 138
Figure 6-5: Farm size and efficiency (with LOESS lines drawn in) 139
Figure 6-6: Hours off-farm and efficiency (with LOESS lines drawn in) 139

Introduction

This study focuses on the livelihood strategies of farmers and farm households in three regions of rural Peru – the sierra, or mountain region, the coastal region and the eastern rainforest region. A number of econometric models are estimated using data from a single dataset – the World Bank Living Standards Measurement Survey for 1994. The estimated models relate to the following four topics: (i) the appropriate utility model for farm households in rural Peru, (ii) the intra-familial aspects of the off-farm participation and labour supply decisions, (iii) separability in labour allocation when working off the farm or when hiring in and (iv) the measurement of efficiency in the Peruvian sierra, and links between different forms of diversification (on and off the farm) and efficiency.

The results of all four sets of models tend to complement each other and reverberate against each other in ways that justify the choice of a single data set for all the estimations. The overall picture that emerges is one where the economic lives of households and individuals with and without agricultural assets, in more accessible and in more remote areas, are at least partially illuminated by economic theory and the tools of econometrics.

Background

The study adds to the body of research that has recently begun to accumulate on on-farm and off-farm labour supply by rural households in developing and middle income countries. The general topic of income strategies of rural households is interesting for a variety of reasons, both practical and theoretical. In practical policy terms, research in this area is concerned with poverty alleviation and inequality reduction, planning regional and rural growth, reducing migration to cities, measuring efficiency of farms and farm households. With regard to economic theory, the area is interesting because of innovations in modelling market failures or in modelling actions at the individual and at the household level.

Under the general heading of rural households' livelihoods and welfare, a full issue of the journal World Development was devoted to the topic in March 2001 and another, of the journal Food Policy, in August 2001. Literature survey articles and books (Reardon *et al.* 2001, Barrett *et al.* 2001, Ellis, 2001, Lanjouw and Lanjouw, 2001) have also discussed the broader relationship between different on and off-farm work patterns and patterns of national and regional development.

One reason for the 'boom' in research in this area has been a shift in focus by the World Bank and other international agencies towards the household as a unit for observation and study. The bank has been largely responsible for the recent availability of household survey data (mostly since the mid to late 1990s), both on the internet, and from national, international and regional agencies. The availability and consequent use of this data, in turn, has encouraged the development of more finely tuned surveys to deal with specific theoretical or practical issues, such as particular kinds of transaction costs that might induce market failures or particular cultural or institutional features that might modulate decision making.

Another reason for the increase in research in the specific area of household income strategies has been the development of micro-econometric models allied with the computing power to solve them. This enables the analysis of such household level data to be carried out with a greater level of ease, accuracy and reliability than heretofore. Better survey and model design have enabled the testing of specific hypotheses concerning time allocation, transaction costs and market imperfections, underlying utility functions, risk management, institutional demand and many other topics connected to the farm household.

Research objective

The primary objective of this study is to improve understanding of labour allocation decisions by individuals in rural farm households in a middle income country. More specifically, by estimating different but connecting models using the same dataset, it is hoped to better understand how a variety of features of farm households (e.g. household composition, asset position, education levels) relate to a range of intertwined issues concerning household livelihood strategies. These issues include off-farm participation in different sectors, constraints in off-farm and on-farm choices, determinants of farm efficiency and long-term household welfare.

And while the primary focus is on understanding these issues, rather than on policy per se, a number of questions that are naturally important for policy are addressed at different points e.g. the most efficient farm size, gender differentiation in off-farm and on-farm activities, ethnic differentiation in intra-familial interactions. A number of methodological issues also arise e.g. how to test for particular utility models in households with more than two decision makers or whether to use individual or household level data in looking for the roots of non-separabilities.

Research hypotheses

In the initial research proposal for this study the main hypotheses to be tested concerned the potential role of off-farm labour in improving on-farm output allocative efficiency. Allocative efficiency may be difficult or impossible to achieve in middle income or developing countries if there are market imperfections. If there are imperfections, for example, in the consumption goods market then farmers may find it necessary to grow certain low-return crops because of the unavailability of basic foodstuffs. Alternatively, if insurance is not available to risk averse farmers (with good reason to be risk averse), then what appears to be allocative inefficiency may in fact be an efficient response in the management of risk. In this kind of situation, if the hypothesis that doing off-farm work increases allocative efficiency was found to be true (because of off-farm work providing insurance against certain kinds of crop failure), this would be yet another reason to focus rural policy makers' attention on off-farm work and away from what has often been a uni-focal view on agriculture.

As the literature review developed, other hypotheses concerning the nature of family interactions and constraints came to the fore. How do families reach decisions? Does the household act as a single unit, or as a group of individuals engaged in a particular bargaining game with each other? How can one model these processes, and what is the best way to think about them empirically? The Collective model (Chiappori, 1988) seems to provide an interesting way to think about these things. In essence, the argument is that for most empirical work we do not need to think about exactly how decisions are arrived at, but only about certain properties of the final arrangement. If these properties (mainly properties associated with Pareto Efficiency) can be shown to exist, then deductions can be made about the power structure within the family and the likely individual reactions to changes in exogenous variables. Testing for the Collective model in a rural setting, where there are often more than two adults in a house, and where there are different cultures, with different modes of intra-familial interaction, is something new. Hypotheses concerning the appropriateness or otherwise of the Collective model gave a reason to focus on individuals within the household, and not just the household as a unit.

The third set of hypotheses to test concerned the benefits or otherwise of focusing on individuals rather than, or as well as, the household. Would there be statistical benefits, in terms of improved model fits, and would there be substantive benefits in terms of knowledge gained about intra-familial behaviour? These would be difficult to predict in advance, but it was likely that there would be some. It was also quite likely that focusing on the individual could help improve some tests designed to check that the separable version of the farm household model was appropriate. Since decisions seem to be made at least partly at the individual level, it was felt that separability of off-farm labour allocations should also be tested at the individual level, to the extent that that is possible. The issue of separability had come to seem important because it is closely related to the imposition of constraints on choice and like allocative inefficiency was related to the whole question of imperfect markets which had initially motivated the research.

Justification for the research

The Living Standard Measurement Surveys, begun in the mid 1980s, are fairly complete data sets. They include data on prices, community level variables, households level variables, detailed consumption data, reasonably detailed farm level data, both on outputs and inputs, and data on individuals, their health, their education, their work on and off the farm. This seemed to be an under-used resource where, because different models could be estimated from the same data, a cross fertilization might occur, and the results from one model help clarify or lead to interesting hypotheses for another. The fact that the same set of variables could be refracted through different models seemed like a valuable opportunity that had been available for a long time but not yet taken up.

Each of the four main sets of models in Chapters 3-6 also has their own justification.

Firstly, deciding on the appropriate utility model for farm households in rural Peru is important theoretically, and for policy. The test in Chapter 3 of whether the more common Unitary utility model or ‘general’ form of the Collective Utility model is appropriate for families with more than 2 adults is new in the literature. It is at least potentially useful to do this because knowing which utility model is appropriate in any given situation may be important. Different utility models can lead to different policy implications. Based on some remarks in the anthropological literature, utility models are also tested for different linguistic groups and different degrees of households' market embeddedness. Although this is quite simple to do, it has not been done before.

Secondly, the intra-familial aspects of the off-farm participation and labour supply decisions are important because of what is revealed about the behaviour of operators, spouses and family members. The literature is full of models examining the behaviour of the household as a whole, or the operator, or the operator and spouse, but there are few studies which examine the behaviour of all adult family members.

Thirdly, separability in labour allocation when working off the farm or when hiring in may seem like a fairly arcane topic. But separability in labour allocation decisions is important for a number of reasons. If consumption and production are not decisions that are made separately, then the cross effects must always be taken into account by policymakers. Also, particular ways in which separability breaks down can have quite important consequences. For example, if monitoring costs on hired labour increase with the amount of labour, this has strong implications for the most efficient farm size distribution. In the separability chapter (Chapter 5) a simple new test is proposed to help diagnose the causes of non-separability and a relatively new technique is used for the first time on individual level data, rather than household level data.

Finally, in Chapter 6, the measurement of allocative, technical and scale efficiency in the Peruvian sierra, and the links between these and different forms of diversification (on and off the farm) provide further insight into why choices are made, and link off-farm choices and allocative efficiency in ways that are new in the literature, even if the techniques for measuring efficiency are standard.

Theoretical framework and data

The theoretical framework for the work comes from the farm household model, the genesis of which has been described by Taylor and Adelman. (2002). The farm household model provides a very flexible framework for analyzing a number of issues connected to the farm household. These include production and consumption issues, labour deployment, capital investment, risk management, human capital investment, dynamic optimisation, separability in decision making and many other issues. The household is modelled as maximizing a utility function, whose main arguments are usually consumption and leisure. The framework is so flexible that sometimes risk aversion, non-separabilities, or simple desire to be on the land (Lopez, 1984), can lead to a divergence between utility and profit maximisation, and behaviour that may seem inefficient can be accounted for.

The farm household model generally assumes that households have what is called a unitary utility function. The household is the unit of analysis, and intra-household conflict is assumed away, whether because of their being a dominant person within the family or because of neat aggregation of individual preferences to a single household welfare function, or for some other reason. There is no reason why the farm household model should not be compatible with a Collective model, which assumes that some form of intra-familial bargaining takes place and that power can matter. So the overall theoretical framework for the thesis is the farm household model in its Collective guise. This form of the farm household model is open to criticism (Huffman, 2004) but it accords with one aim of the study which is to try to focus as much on what goes on within the household as on what the household does or does not do as a unit.

Other family models, less in the neo-classical tradition, are available, and are interesting both in themselves and in their implications. Most of the conclusions reached in the anthropological literature, for example, concerning family interaction in indigenous communities in Peru, and the on-farm division of gender roles, have also been reproduced in the econometric analysis in this study. The richness of the case study approach cannot be attained by econometric modelling; but the more generalisable results from quantitative models can confer authority on case studies.

Limitations

There are many limitations both on what a study like this can accomplish and on what this particular study has accomplished. In the first instance, it is completely within the neo-classical framework. The utility maximizing household of neo-classical economics, whether Unitary or Collective, is an asocial entity (at least, outside the household), not a flesh and blood fully differentiated unit interacting with others. Of course, the universality of neo-classical economics (to the degree that it is not totally wrong) is also its strength, in that it provides a framework for analysis of large numbers of people at the same time, and goes some way towards enabling welfare analysis to be possible. But the insight provided by some sociological or institutional approaches (Bettio and Villa, 1998; Folbre, 1994) to family interactions is inevitably lost.

The actual tools for the analysis, econometrics and some linear programming, are generally seen as part of the neo-classical 'package'. There are only numbers, no names or faces, in this study. This leaves, arguably, a gap between what the model results say and what a policymaker might be wise to take from them. This gap could be filled by local knowledge, a feel for what is driving results, a sense of what is missing. That the gap is there may be a disadvantage when trying to think about household strategies and intrafamilial interactions, which after all have a universal aspect to them. It is potentially an even bigger disadvantage when the focus is on day to day local features that can cause market imperfections, such as monitoring costs of hired labour, the costs of getting to the best paying market, and getting information about it, or the influence of language and culture on decision making. Reading a portion of the sociological and anthropological literature about the area, especially the sierra, will, hopefully, have helped to bridge it, but it does potentially limit the validity of some of the interpretations in unforeseeable ways.

Known limitations of what the models can and cannot (and do and do not) do are referred to as they arise, and most especially in section 7.3 of the concluding chapter, where issues for further research are discussed.

Overview (by chapter)

In Chapter 1, the standard farm household model is briefly introduced and the main results of research into off-farm labour supply are outlined for developing and developed countries. The Collective model is introduced at the end of the chapter. In Chapter 2, the Peruvian dataset (for 1994) is described and explored along with some background information on Peru. In Chapter 3, the Collective model of household decision making is tested for households with 2, 3 and more adults, with results suggesting that in examining household decisions it is important to look at the individual as well as at the household as a whole. In Chapter 4, various simple reduced form models of the off-farm participation and hours decisions for individuals are presented and analyzed. The progressive disaggregation of the models reveals new results on intra-household accommodations. In Chapter 5 the literature on labour market separability is briefly reviewed, and separability is tested at the household level, but also at the individual level. The results suggest that individual level tests are an improvement, at least in some cases, on household level tests, and that off-farm work is rationed and hired in work is subject to non-linear transaction costs. Chapter 6 reverts to the household level for an examination of the effects of off-farm participation on agricultural technical, allocative and scale efficiency, measured using Data Envelopment Analysis (the literature on which is very briefly reviewed), and long-term household welfare (as measured by total household expenditure). Conclusions are drawn together and implications discussed in Chapter 7.

Summary

This thesis is a study of farm household choices, constraints and welfare in rural Peru, for 1994. The study incorporates literature reviews of farm household models and the main results concerning off-farm participation for developed and developing countries. There is also a review of the Collective utility model, as well as much shorter reviews of the literatures on separability in labour allocation and on Data Envelopment Analysis. The data used is from the 1994 Living Standards Measurement survey, conducted by the World Bank. 4 sets of econometric or programming models are estimated, using a sample of about 1000 farm households. These models are (i) an Almost Ideal Demand System, which is used to conduct tests of the Collective versus the Unitary utility models for families of different sizes and for different linguistic groups, (ii) a set of mainly multinomial logit based off-farm participation models, (iii) models of on-farm labour time designed to test for separability in working off-farm and in hiring in and (iv) a Data Envelopment Analysis which is designed to measure allocative, technical and scale efficiencies in the Peruvian sierra, and to examine the factors affecting the different kinds of efficiency. The results are analysed at the end of each chapter, and then connections are made in a final chapter.

The 10 main substantive results are:

(i) the Collective model is probably appropriate generally for rural Peru for both 2 and 3 adult families; there are suggestions, though that the intra-household

Pareto efficiency it implies is more likely to exist for non-Spanish than Spanish speaking families;

(ii) a farm size distribution with a mode of around 2 hectares is the most technically efficient,

(iii) the two main ways of getting out of poverty appear to be to have farm assets or to increase education – the former seems more important for the extremely poor, the latter for the slightly better off;

(iv) on-farm crop diversification is strongly negatively correlated with off-farm work participation, suggesting that these are alternative risk management strategies; crop diversification is positively associated with improved technical efficiency and negatively with allocative efficiency; crop diversification decisions are mainly handled by the males in the household;

(v) market embeddedness (defined as the proportion of crops sold) is negatively correlated with off-farm work participation, and positively and strongly correlated with all forms of efficiency and income; cash flow in the household, whether it comes from being embedded in the market or from off-farm work, is handled by all family members;

(vi) higher education (as measured by years of schooling) is associated with greater off-farm participation for the better off in non-agricultural wage labour (or in self-employment, in the case of female spouses) and with higher off-farm earnings; higher education reduces the probability of being rationed in off-farm work; it is associated with greater on-farm efficiency (but only for those who are very highly educated); the returns to education in off-farm work generally outweigh the returns in on-farm work;

(vii) being non-Spanish speaking increases the chances of working on one's own farm unpaid (and being involved in unpaid labour exchange); family interaction is likely to conform to the Collective model; one is much more likely to be in the lowest permanent income quintile, and to be rationed in off-farm work hours; efficiency of all kinds is lower for non-Spanish speakers, but it is difficult to distinguish a cause – market emdeddedness and remoteness variables seem to be more important than language grouping per se in explaining efficiency;

(viii) women are less likely to work off-farm than men, and less likely to work for wages than to be self-employed; they appear to have a higher shadow value of home time irrespective of the number of children, and households with relatively more females are more likely to hire in; there is little noticeable difference in on and off-farm returns for female labour compared to male labour (though female wages are generally lower, they work less often in the wage sector); off-farm, women are more likely to work in crafts and sales or young women in the professions (teaching mainly), while men work off-farm for other farmers or in construction or transport,

(ix) a variety of other factors determine off-farm participation, including age, on-farm experience, local demand for non– and off-farm labour, and the actual main crops farmed (rice farming households, for example rarely work off-farm, while anyone from households whose main crop is maize is liable to, and spouses on banana farms often do); the off-farm share of total earned income declines with farm size and with permanent income (as measured by per capita expenditure); shares increase with education, but not linearly, and the most educated quintile have a lower share of off-farm income (but substantially higher off-farm earnings) than the second most educated quintile; having off-farm work does appear to be associated with allocative efficiency, at least for those households working more than 1500 hours off-farm, but the relationship is not very strong;

(x) between approximately 25% and 45% of those who work off-farm for wages are constrained in the amount of work they can get at the going wage; these are mainly lower paid workers in areas such as skilled construction labour and farm labour. The difference in education levels between the constrained and unconstrained is very stark (8.15 years, on average, for the unconstrained compared to 5.14 years for the constrained).

There are also four methodological innovations in the study. These are: (i) a test for the three-decision-maker version of the Collective model; (ii) a test to distinguish rationing from transaction costs as determinants of non-separability, (iii) the use of individual data in participation and separability estimations, (iv) the use of Censored Least Absolute Deviation (CLAD) models to test for the link between allocative efficiency and off-farm hours or earnings.

1

The farm household model

Introduction

In this chapter the main theoretical framework for modelling farm household decisions is presented. First, the Unitary Utility model is presented for households as a single unit and then for two-adult farm households. Some labour allocation models and the main results for developing countries are then summarized. Results for developed countries are also summarized, and the two sets of results compared. Some of the policy implications of these studies for developing countries are discussed, and the current research agenda is outlined. In the second part of the chapter, the Collective Utility model is presented in a general context (not specifically for farm households[1]).

1.1 The unitary farm household model – introduction

The deployment of farm household labour on and off the farm in middle income and developing countries has been the subject of intensive research by economists since at least the early 1990s. This burst of activity has been made possible by the large number of national and regional household surveys undertaken in developing countries in the late 1980s and throughout the 1990s, often under the aegis of the World Bank or following the format of the World Bank's Living Standards Measurement Surveys. These surveys provide sufficient information to be able to calculate nearly all the relevant descriptive statistics regarding off-farm work and to model a large variety of participation, labour supply and income share equations.

The initial objective of much of this research has been simply to find out how much household income comes from off-farm or non-agricultural sources and to measure how much time is devoted to these activities, and in what sectors they take place. Thanks to the number of recent surveys undertaken the answers to these questions are

[1] A brief discussion of the Collective Model in a Farm Household context is relegated to the Appendix. This is because the use that is made of the Collective Model, in Chapter 3, does not depend on the object of study being farm households.

now largely known. Beyond these descriptive questions, some of the more important issues examined in the recent research include the determinants of the off-farm work decision itself , possible barriers to working off the farm and access to different types of off-farm work (Savadogo *et al.* 1998; Reardon *et al.* 1994; Were, 1996; Dercon, 1998; Dercon and Krishnan, 1996). The roles of local and more distant labour markets have also been examined (Stark, 1991; Rozelle *et al.* 1999; Taylor and Adelman, 2002; Becker *et al.* 2003), as have questions relating to the dynamic effects of the diversification of family income sources on household income growth over time (Barrett *et al.* 2001, Block and Webb, 2001; Canagarajah *et al.* 2001). The main results of this research are summarized in sections 1.1.6 and 1.1.7 below and some of the main topics on the present research agenda are listed in section 1.1.9. Two considerations in particular suffice to motivate this ongoing research.

Firstly, off-farm work is important because it makes up a large though varying part of family incomes for many farm households, both in the developing and developed world. Table 1-1 summarizes information from Reardon, *et al.* (2001) and Barrett and Reardon (2000). From this Table both the general importance and the variation in importance of non-agricultural incomes across Latin America and Africa is clear:

Table 1-1: Percentage of income of the rurally employed not from agriculture

	Year	%			Year	%
Botswana	1974/75	54		Brazil	1997	39
Botswana	1985/86	77		Chile	1997	41
Burkina Faso	1978/79	22		Columbia	1997	50
Burkina Faso	1981/84	40		Costa Rica	1989	59
Ethiopia	1989/90	36		Ecuador	1995	41
Gambia	1985/86	23		El Salvador	1995	38
Kenya (central)	1984/85	42		Haiti	1996	68
Kenya (western)	1987/89	80		Honduras	1997	22
Lesotho	1976	78		Mexico	1997	55
Malawi	1990/91	34		Nicaragua	1998	42
Mali	1988/89	59		Panama	1997	50
				Peru	1997	50

In both continents the shares tend to be a lot higher than the share accruing from wages IN agriculture (e.g. even in Honduras, which has a low share of non-agricultural income, the ratio of non-agricultural income to agricultural wages among the rurally employed is 1.3 to 1). Together, both income sources make up a large and growing part of rural incomes. This is motivation enough to study non-agricultural incomes and farm wage work and, through this, to search for answers to such questions as which impediments if any hinder entry to better paying off-farm sectors, how important off-farm income is for family welfare, what the relationship is between different off-farm work and on-farm practices.

From a policy point of view, these questions are clearly important. The measurement and alleviation of poverty, social welfare and aid programmes, institutional development, regional development and many other types of intervention will almost certainly be affected by an awareness of the changing nature and effects of different types of rural income generation. To take one example, there is often a bias towards agriculture in rural development, or – if not – then towards manufacturing to increase non-agricultural employment in rural areas, (Reardon *et al.* 2001). In most cases, however, the majority of rural employment is actually in the service sector, not manufacturing, and it is also the service sector which is growing fastest in most countries. Training for this sector, or other relevant policies, may be fruitful development options.

Secondly, apart from its general importance for development and poverty alleviation, the decision to work off-farm is an important one for most individual farm family members and, in this regard, an interesting question is how decisions concerning individual and family income generating and sharing strategies are arrived at and executed, given available options and the existing constraints. Decisions within the farm family concerning who should work at what activities and for how long are decisions that may be made differently according to class, region, religion, custom etc., with often unintended consequences for development. Appropriate ways of modelling the decision making process are essential, therefore: in some cases, it may be appropriate to treat all farm families alike as undifferentiated utility or profit maximizing units, but in many other cases, perhaps most, a more differentiated approach may be beneficial. Understanding the determinants of and possible barriers to off-farm work may be a more layered process than it first seems. The importance of certain determinants and constraints can be easily missed or misjudged if an undifferentiated approach is taken (e.g. the on-farm consequences of increasing off-farm work for men may differ from those resulting from increased off-farm opportunities for women or adult children; both of these may in turn be affected if older generations live in the family home and play a role in family income generation or home production).

The motivation of exploring the link between intra-familial decision making processes and development is a strong one, since ignoring such processes may lead both to reductions in general family welfare and to reductions in welfare of different types of family members. Moreover, the possible understanding gained of intra-familial decision-making in the area of off-farm work may be applicable to other areas (e.g. child nutrition, development of human capital).

1.1.1 The farm household model – specification and estimation

Before continuing, a few basic terms are now defined: i) the term 'off-farm' work, in this thesis, means all work by the household which could increase household income, apart from family agricultural work on that portion of land for which the household has the right to the bulk of residual income; ii) 'non-agricultural' work means all remunerable work by household members that is not in the agricultural sector; iii) the term 'diversification' will signify crop diversification on the farm and finally, iv) households physically based in a rural area will be the working definition of a rural household.

In each of these four cases, more precision will be used when required.

Theoretical and econometric analysis of the determinants of off-farm work, and farm household labour supply in general, usually takes place implicitly or explicitly within the framework of a farm household model (Singh, Squire and Strauss, 1986; Taylor and Adelman, 2002). Often, the unit of analysis is either the household as a whole, or the farm operator. For the explicit analysis of multiple person households the usual assumption is of a joint or Unitary utility function, which can be justified (Samuelson, 1956) by consensus on the 'ethical worth' of the welfare of family members or (Becker, 1981) by the existence of a household 'head' who cares about the welfare of the other members.

An alternative framework for analysis is the Collective utility model (Chiappori, 1988), which generalises the co-operative game theoretic household bargaining models of Manser and Brown (1980) and McElroy and Horney (1981), and which has gained in popularity since tests in developing countries began to find strong evidence against the unitary model (Strauss and Thomas (1995) comprehensively review the evidence). Within the collective framework household utility is a weighted average of the utility of its members, the weights corresponding to the members' bargaining strength. If a particular utility function is specified, the 'sharing rule' can be recovered through estimation. The appropriate utility model for rural Peru is tested for in Chapter 3.

1.1.2 Unitary household model – farm household as a unit

The decision-making unit chosen will inevitably affect econometric estimation strategies. In the standard form (household as an undifferentiated unit), the usual specification is as follows: the household is assumed to maximize a one-period utility function whose arguments are household consumption and leisure and the main constraints on which are the time endowment, a farm production constraint and income.

$$\max_{T_h, C, T_f, T_m} U(T_h, C; Z_h) \qquad (1-1)$$

subject to

$$T = T_h + T_f + T_m \qquad (1-2)$$

$$C = g(T_{f;}p, Z_f, H_f) + w_m(H_m, Z_m)T_m + V \qquad (1-3)$$

$$T_m \geq 0, \qquad (1-4)$$

where T_h = home time, or leisure

C = consumption of goods other than home time

Z_h = household characteristics that affect preferences

T = time endowment

T_f = own farm work time

T_m = off-farm work time

g = farm income function

p = vector of prices of agricultural outputs and inputs, excluding family on-farm labour

Z_f = fixed farm input

w_m = wage rate for off-farm work

H_m = human capital which influences wage level

H_f = human capital which influences farm output

Z_m = other factors affecting the wage level

V = non-labour income.

The maximization problem in this specification can be characterised as a two-step process (Huffmann and Lange, 1989). First, farm profits are maximized and then household utility is maximized. Utility maximization may involve extra work off the farm or it may not. Whether or not it does so depends, in a 'frictionless' world, on whether available wages are at least as high as reservation wages (w_h). If available wages are higher than reservation wages, then both separability and recursivity hold, and i) consumption decisions are separate from production decisions and unaffected by them (except through their effects on full income) and ii) the model can be estimated recursively (joint estimation, for example, of both farm profits and household labour supply is not necessary).

This, simple, version of the model assumes, for example, that hiring-in of labour does not occur. However, total wages paid for hired labour could be added to the model as a negative amount on the right hand side of Equation 1–3 if hiring in is being modelled. Later in this thesis, in Chapter 5, hiring-in is explicitly added to choice set of households (up until then the issue is ignored, with the focus mainly on off-farm participation), and is formally incorporated into the farm household model in Equations 5–1 to 5–3.

Clear comparative static predictions can be derived analytically from the standard model but, first, it is useful to define the 'participation function': i:

$T_m > 0$ if: $i^*(H_m, Z_m, H_f, p, Z_f, Z_h, T, V) \equiv w_m(H_m, Z_m) - w_h(H_f, p, Z_f, Z_h, T, V) > 0$

but, on the other hand, $T_m = 0$ if

$i^*(H_m, Z_m, H_f, p, Z_f, Z_h, T, V) \equiv w_m(H_m, Z_m) - w_h(H_f, p, Z_f, Z_h, T, V) \leq 0 \quad (1-5).$

Lee (1998, Chapter 4) draws the following comparative static implications from the standard model, where the symbols in each box indicate the direction of the effect of an increase in the variable in the left hand column:

Table 1-2: Comparative statics in the unitary utility model

	Wage	Shadow Wage at Zero Off-Farm Hours	i* = Wage – Shadow Wage
Human Capital for Off-Farm Work (H_m)	+	0	+
Favourable Off-Farm Labour Market Situation (Z_m)	+	0	+
Time Endowment (T)	0	-	+
Unearned Income (V)	0	+	-
Preference Change in Favour of Home Time (Z_h)	0	+	-

Farm Output Price (if labour is a normal input) (p)	0	+	-
Farm Input Price (complementary to labour) (p)	0	-	+
Fixed Input (complementary to labour) e.g. farm size (Z_f)	0	+	-

This picture is made substantially more complicated if off-farm wages are non-linear in time worked, which is likely to be the case if off-farm work is in self-employment rather than in paid employment. The same complications arise if there is any feature of off-farm work (e.g. non-linear transaction costs) that leads to non-linearity in effective returns. In both of these cases, and indeed any case where returns are non-linear, separability and recursivity do not hold. Non-linear returns leads to simultaneity in the production and consumption decisions[2].

1.1.3 Unitary farm household model with 2 (or more) members

When the farm household is broadened out to two or more people per household then the maximization problem within the unitary utility framework changes to the following:

$$\underset{T_{h_i},T_{m_i},T_{f_i},C}{Max}\ U(C,T_{h_1},T_{h_2},....,T_{h_n};Z_{h_1},Z_{h_2},Z_h) \qquad (1-6)$$

subject to

$$C = g\,(p;T_{f_1},T_{f_2},....,T_{f_n};Z_f)+\sum_{i=1}^{n} w_{m_i}T_{m_i}+V \qquad (1-7)$$

$$T_i = T_{h_i}+T_{f_i}+T_{m_i}, i=1,2,...,n \qquad (1-8)$$

$$T_{m_i}\geq 0 \qquad (1-9),$$

where i is an individual subscript and n is the number of people in each household who jointly optimise (usually two). Kuhn-Tucker first order conditions are obtained from the above, and participation conditions for the operator and spouse (the model has only to my knowledge, been explicitly estimated for more than the two person household in the three cases reported section 1.1.5) can be obtained directly from these:

$$\text{PC1: } w_{m1}-w_{h1}\leq 0, T_{m1}\geq 0, (w_{m1}-w_{h1})T_{m1}=0$$

$$\text{PC2: } w_{m2}-w_{h2}\leq 0, T_{m2}\geq 0, (w_{m2}-w_{h2})T_{m2}=0 \qquad (1-10),$$

[2] The model as presented does not include the hiring in decision, discussed in chapter 5. But the same holds true: non-linear returns to hiring in results in non-separability.

where w_h, again, is the reservation wage, subscripted for each person.

There are normally two ways in which these participation conditions are related to i*, the participation function:

i) directly, through a multivariate approach based on the following econometric system –

$$i_1 = 1 \text{ if } i_1^* = \beta_1' x + e_1 > 0$$

$$i_1 = 0 \text{ if } i_1^* = \beta_1' x + e_1 \leq 0$$

$$i_2 = 1 \text{ if } i_2^* = \beta_2' x + e_2 > 0$$

$$i_2 = 0 \text{ if } i_2^* = \beta_2' x + e_2 \leq 0 \qquad (1-11),$$

where x is the vector of all the exogenous variables from the maximization problem.

If both error terms are normally distributed then this system can be estimated using a bivariate probit model.

Or ii), using an indirect utility approach:

$$U_{ij}^* = \beta_j' x_i + e_{ij}, \qquad (1-12),$$

where U* is the maximized household utility from (1 – 6) and j indexes the four possible choices facing the household: no off-farm work, only the husband has off-farm work, only the wife has off-farm work, both persons have off-farm work. Restrictions may be imposed on the coefficients of variables which affect one or other partner only. If the error term follows an extreme value distribution, the joint choice process can be modelled as a multinomial logit choice model.

There are advantages and disadvantages to both approaches. The bivariate probit approach has a theoretical flaw in that the simple participation function delineated in (1 – 5) now also includes the actual wage of the partner as an argument in each person's shadow wage function, as well as the relevant exogenous variables for both partners and the other exogenous variables. However, in some cases the partner may not actually work off-farm. It is impossible to know in advance whether results of a reduced form participation model (containing some couples where both work off-farm, some where only one does, and some where none do) will be strongly affected or not if this is the case, but Brick *et al.* (2005) suggests they may be.

The main problem with the multinomial version is more practical. A multinomial probit model would allow both for correlations among the random variables, and for heteroskedasticity, but convergence of multinomial probits is extremely difficult to achieve (Keane, 1992). While convergence is not a problem for the multinomial logit model, the assumption of identical, independent errors may be too strong. And, at the very least, the independence of irrelevant alternatives should be tested for.

The bivariate probit approach is far more common in the literature (Huffman and Lange, 189; Gould and Saupe, 1989; Tokle and Huffman, 1991; Kimhi, 1994b; Olfert *et al.* 1992; and in Ireland, Keeney, 2000). It has been extended or adapted to allow in particular for simultaneity in decision making (Findeis and Lass (1992) who allow for simultaneity in the hiring and participation decisions)) and for the application of panel methods (Lass and Gempesaw, 1992). Ahitov and Kimhi (2002) use a multinomial approach (in a dynamic, simultaneous setting), as do Kooreman and Kapteyn (1987) and Brick *et al.* (2005).

Whether one approach or other is used, the ensuing step, if one moves beyond modelling the reduced form participation decision, is usually to estimate operators' and spouses' wage and hours equations (off-farm labour demand and off-farm supply). Selection issues and possible endogeneity of earnings, income and wealth variables are common problems to be overcome in the latter estimations. More structural models are rare in the literature. Lundberg (1988), using a conditional supply approach, and Kimhi and Lee (1996), using minimum distance estimators, are two exceptions. Approaches with more than two decision makers have not been attempted for developed countries.

1.1.4 Home production and unpaid farm work

The farm household model has been extended (Jacoby, 1995; Apps and Rees, 1997), as have household models in general, to include the distinction between home production and leisure. The theoretical issues surrounding this adaptation of the model have been discussed in Pitt and Rosenzweig (1986), Apps and Rees (1997), Chiappori (1988) and in Apps (2003). The inclusion of home production within the farm household framework affects the theoretical conclusions of both the multi-person unitary model (the Slutsky condition is altered) and the Collective model (stronger assumptions concerning technology and preferences need to be made for the 'sharing rule' to be recoverable). The inclusion of home production is compatible with a 'traditional family model', where the male operator chooses his labour supply independent from that chosen by his spouse but the converse is not true (Lundberg, 1988; Bowen and Finnegan, 1969; Hall, 1973; Heckman and McCurdy, 1980; Hausman, 1980, 1981; McCurdy, 1981). It is also compatible with some non-neo-classical models of household behaviour.

For the purposes of econometric estimation of labour supply, the effect of including home production as a separate category of work may be to complicate estimation, since hours of farm work and home work are both simultaneously decided, and recursivity will not hold even if there is no leisure. The main problem, though, with incorporating home production is probably with believing in the data. Detailed time diaries are almost certainly needed before home production hours can be plausibly included in labour supply models.

Unpaid farm work by adult family members is a category not usually examined in farm household models. With regard to on-farm work, it is assumed, simply, that – if they work on the farm – each (type of) family member does so until marginal revenue (the marginal value of an extra unit of that type of family farm labour) equals marginal cost (the shadow value of that type of family member's leisure time and/or home production). If they do not work at all on the farm it is assumed that marginal revenue is always less than marginal cost.

In the case where they do work on-farm, it is a separate question as to how family on-farm returns are distributed, no matter in whose name they are declared. In households where unpaid farm work by adults is common, it could be, therefore, that the basic assumptions of the farm household model, either in its traditional, unitary or collective variant, are being met; or it could be that some other non-neo-classical model of family decision making is needed.

In the results of Chapters 3 and 4 some differences are found between people and household types regarding who declares farm returns.

1.1.5 Modelling the determinants of off-farm participation and income in developing countries

For developing countries, the seminal works using the neo-classical treatment of peasant labour supply (the 'farm household model') are by Rosenzweig (1980) and by Barnum and Squire (1979). These build on the pioneering work of Lau *et al.* (1978) and Kuroda and Yotopoulos (1978), in which the farm household model was first developed (initially, to explain the empirical puzzle of falling supply of rice following on from a policy induced increase in price)[3], and on the family labour models of Ashenfelter and Heckman (1974). Singh *et al.* (1986) summarize much of the early theoretical and empirical work using (unitary) farm household models in developing countries. The separability assumption is held in all of these models (marginal conditions are determined by market wages), and models are estimated recursively. The usual assumptions yielding a recursive model (linearity of wages, zero or linear transaction costs in labour markets, perfect substitutability of family and hired labour) were dropped by Jacoby (1993) on foot of evidence (e.g. Deolalikar and Vijverberg (1987) on hired labour) against them. Jacoby's model uses Peruvian data and builds on both Lopez' (1984) paper, using Canadian data, where for the first time the implications of dropping recursivity are modelled, and Benjamin's (1992) paper, which develops a series of tests of the separability of labour. The main problem that Jacoby deals with is the following: much off-farm labour in developing countries is in self-employment, leading to non-linearities and endogeneity of off-farm wages (and, hence, non-separability). Jacoby's solution is to estimate a farm production function and to use the individually fitted marginal products of different kinds of family labour (male, female) as shadow wages for that kind of labour. These shadow wages are then used in non-recursive labour supply models of men's and women's total labour supply (on-farm, off-farm and home production activities). Jacoby finds that peasant allocation of time in rural Peru is consistent with an allocation which maximizes a family utility function. It can therefore be modelled within a simple unitary framework. Skoufias (1994) extends Jacoby's model to India, while Abduali and Regmi (2000) apply it to Nepalese data. The potential importance of the separability assumption is seen at its clearest in the Nepalese paper, where results are contrasted

[3] The 'solution' to the puzzle was that the income effect of the price increase led to greater consumption of home grown rice, to the extent that less rice was made available for sale on the market.

for alternative specifications based on assumptions of separability and non-separability. If separability is accepted the labour supply curve is backward bending for both men and women: if it is not, then the estimations show that labour supply increases as shadow wages rise.

Newman and Gertler (1994), using the same dataset as Jacoby (from the 1985 Peruvian Living Standards Measurement Survey), take an altogether different approach. They allow for interdependence in family decision-making and they refrain from using farm income and production information (which they do not trust), and so do not estimate shadow wages directly. The family jointly determine family consumption and the labour supply / leisure of individual members. A three equation system is simultaneously estimated: i) a wages equation, ii) a marginal returns to farm work equation and iii) a marginal rate of substitution (MRS) equation. The Kuhn Tucker conditions from the multi-person maximization problem ((1 – 6) – (1 – 9)) are exploited to recover missing information for the latter two equations.

The economy of information needed makes this an attractive model, but it is complicated by the intra-household interactions (of all adults – not just the farm couple, as in most other multi-person models), which lead to a likelihood function that is close to intractable. Newman and Gertler opt for the simplest possible assumptions about covariances (i.e. zero covariance) between the errors of family members in their three equations to ensure convergence. In the reported results all restrictions on the influence of intra-family interactions are comprehensively rejected. The effect of lifting these restrictions on the reported coefficients and on the predictive performance of the model is, however, very small. The actual payoff from the complexities involved in formulating a structural model is in the welfare analysis, where the model allows a detailed simulation of the effects of changes in wages or farm returns on time allocation, and through the MRS equation (which enables a monetization of leisure), on family welfare.

Other 'semi' or 'partially'-structural models allowing for non-separability have been estimated by Lazlo (2002), and Pascual and Berbier (2001). Lazlo estimates a labour supply model for Peruvian households and Pascual and Barbier (2001) estimate a labour supply model (incorporating hours of other family members, and of hired labour) for operators and spouses in Mexico.

Two reduced form models (without either wage or family hours variables on the right hand side) that include more than a single intra-familial interaction are by Fafchamps and Quisumbing, and by Malchow-Moller and Svarer. Extending the two-step (family/individual) model of Fafchamps and Quisumbing (2000), Malchow– Møller and Svarer (2002) assume that all adults, not just the operator and spouse, participate in decision making. They use a random effects multinomial specification in their estimation of Nicaraguan off-farm labour supply, where the estimation is at the level of the individual adult, and the random effect is at the level of the household. Their justification for the multinomial logit procedure is not based on the usual random utility arguments. In their formulation, optimised labour supply in different sectors replaces utility as the latent dependent variable:

$$S^*_{ij} = \beta'_j x_i + v_h + e_{ij}, \qquad (1-13),$$

where the x variables are the exogenous variables from the household maximization problem, S* is optimal labour supply of person i in job category j, h is a household subscript attached to a household specific error, and the last term is the error specific to each individual in each of their possible job categories.

The continuous latent dependent variable is in turn replaced operationally by a categorical participation variable (the three categories being participation in agricultural off-farm employment, participation in non-agricultural off-farm employment and – the base category – not participating in off-farm employment). The reason for switching from an hours supplied to a categorical participation dependent variable is the lack of faith they have in the hours reported by Nicaraguan farmers. This switch in framework enables a direct testing of the effects of changes in the independent variables on the labour supply decision, where these changes take place at the level of the individual, other household members, whole household, farm and local economy. The problem of estimating a random effects multinomial logit model is overcome by adapting a procedure used by Chen and Kuo (2000) whereby an SAS command for a nonlinear Poisson count estimation can be adapted so that the likelihood function actually maximized turns out to be substantially the same as that for a random-effects multinomial logit model.

Much of the rest of the literature estimating the determinants of off-farm work is in the form of participation and / or labour supply models within the standard unitary framework – the 'one person' household, or the operator. Very occasionally, the individual, as opposed to the household or the operator, is the only or the main unit of analysis, even if two or more individuals are from the same household (Ruben and van den Berg, 2001; Corral and Reardon, 2001). The household participation estimations tend to use probit or logit models (Corral and Reardon, 2001; Elbers and Lanjouw, 2001; Berdegue *et al.* 2001; Ferreira and Lanjouw, 2001; Adulai and Crole-Rees, 2001[4]; while the supply (hours) and demand (wages) models are rarely estimated (Woldehanna and Oskam, 2001, are an exception: they use a tobit model to estimate a household off-farm hours equation).

Closely related groups of models look at the determinants of the level of off-farm incomes (Lanjouw, 2001; Berdegue *et al.* 2001; Ferreira and Lanjouw, 2001; Corral and Reardon, 2001) and at the determinants of the extent of household diversification (Block and Webb, 2001). A further group of models looks at the determinants of either the different types of off-farm work engaged in by the household (Barrett at al, 2000; Barrett *et al.* 2001; Woldenhanna and Oskam, 2001; Ruben and van den Berg, 2001; deJanvry and Sadoulet, 2001; Lanjouw, 1998, 2001; Corral and Reardon, 2001) or of the income shares accruing from different types of off-farm work (Escobal, 2001).

In much of this work there is an implicit assumption of portfolio management by the farm household (Binswager, 1980; Bar-Shira *et al.* 1997). This is made the explicit focus of investigation in papers testing (and finding evidence for) income smoothing strategies by Reardon, *et al.* (1993) and Ellis (1998, 2001), in turn related to other studies on different aspects of income and consumption smoothing by Rosenzweig (1989), Townsend (1994), Carter (1997) and Paxson (2000), some of which is summarized in Morduch (1995). Deininger and Olinto (2001) regress total household expenditure (as a proxy for permanent income) on the degree of household income source specialization and find a positive relationship, suggesting that income diversification may be a deliberate risk managing strategy. Dercon and Krishman (1996), Barrett (1997) and Reardon (1997) use a variety of models to explore barriers to high-paying off-farm work in Africa and find that wealth and lack of education reduce the possibilities of high-paying off-farm work. Barrett, *et al.* (2001)

---

[4] Ten out of ten models for developed countries reported in Huffman (1988) also use dichotomous probits – although many of these are bivariate probits.

specifically take a diversification approach to the analysis of income strategies throughout the whole of Africa.

1.1.6 Main findings for developing countries

The principal findings of the recent work can be summarized as follows:

Main Descriptive Statistics: rural non-agricultural income makes up roughly 42% of rural household income in Sub-Saharan Africa, 40% in Latin America and 32% in Asia (Reardon *et al.* 2001); in Africa income from non-agricultural self-employment exceeds that from wages, while in Latin America the opposite is true (Reardon *et al.* 2001); rural non-agricultural income is a far greater component of household income than are emigrant remittances (Barrett *et al.* 2001).

Determinants – Pull and Push: the main 'pull' determinant to undertake non-agricultural work is the availability of higher earnings outside agriculture. These are likely to vary from region to region and for different types of people (Reardon *et al.* 2001; Lanjouw and Lanjouw, 2001). The main 'push' factors come about due to missing goods / land / labour / capital and / or insurance markets. Households may be pushed into work off-farm in order, for example, to manage ex ante output and price risk or to deal ex post with agricultural income shocks (Rosenzweig and Stark, 1989; Rose, 1994; Townsend, 1994; Kochar, 1999; Paulson, 2000). They may work off-farm to generate working capital because it is unavailable elsewhere (Savadogo *et al.* 1998; Reardon *et al.* 1994), or because land markets are under-developed and extra land cannot be bought or rented (Barrett and Reardon, 2000). Efstratoglou-Todoulou (1990) develop a test to distinguish the importance of push and pull factors on the decision to diversify. It is necessary to remember, also, that missing markets (e.g. missing credit markets) may at times be an impediment to off-farm work, rather than a push factor (Dercon, 1998).

Preferences: Households may diversify even when no markets are imperfect and when there are no 'pull' factors. This could occur, for example, if preferences were such that the household desires a farming lifestyle even if returns are lower than available off-farm alternatives (Lopez (1984) in a developed country context, assumes differing preferences for on and for off-farm work).

Education and Infrastructure: improved general education (see, inter alia, Corral and Reardon, 2001; Elbers and Lanjouw, 2001; Berdegue *et al.*, 2001; Lanjouw, 2001; Ferreira and Lanjouw, 2001; Adulai and Crole-Rees, 2001; Ruben and van den Berg, 2001; Barrett *et al.* 2001) and improved rural infrastructure (Smith *et al.* 2001;

Lanjouw et al. 2001; Lanjouw, 1998) both tend to increase ceteris paribus non-agricultural participation, hours worked and non-agricultural share of household income, mainly through their effects on wages or returns (though there is at least a theoretical possibility that improved infrastructure will lead to more specialization through increased competition). Education tends to have stronger direct effects on off-farm than on-farm rewards (Lazlo, 2001), but may have indirect effects on on-farm rewards (Yang, 2000).

Gender: off the farm, men tend to work more in paid employment and in better-paying self-employment activities and women in smaller scale self-employment activities (Barrett *et al.* 2001; Woldenhanna and Oskam, 2001; Ruben and van den Berg, 2001; de Janvry and Sadoulet, 2001; Lanjouw, 1998; Lanjouw 2001); women are less likely to be excluded from non-agricultural work in Latin America than in Africa (Reardon *et al.* 2001; Barrett *et al.* 2001); and less likely again in Africa than in Asia (Udry, 1996). In some cases (e.g. Chile, Berdegue *et al.* 2001) women tend to earn more than men in non-agricultural activities, but this is rare.

Children: The results for children are either insignificant (e.g. Lanjouw *et al.* 2001; Abdyuklai and Crole-Res, 2001), or positive for entry to low-productivity non-agricultural work or agricultural wage work (Lanjouw, 2001; Ferreira and Lanjouw, 2001; Ruerd and van den Berg, 2001).

Social Capital: In Tanzania active involvement in village affairs tends to reduce off-farm participation, while increased trust in government officials and membership of clubs/societies etc. tends to increase it (Lanjouw *et al.* 2001); in Mexico, increased ejido social capital is associated with higher household off-farm earnings (Lanjouw, 1998).

Age: age practically always has the familiar positive concave effect, where participation increases until a certain age – usually in the 40s – and then drops.

Credit and Wealth: access to credit and availability of capital often determine access to higher paying non-agricultural work (Dercon, 1998; Dercon and Krishnan, 1996; Escobal, 2001).

Regional Differences: economically wealthier zones may have high rural non-agricultural income shares, irrespective of whether the wealth of the zone is due to agricultural activities or to non-agricultural activities: a lot, in this instance, depends on historical linkages between agriculture and non-agricultural sectors (Reardon *et al.* 2001).

Sectoral Differences: more non-agricultural work tends to be in services than in manufacturing (Reardon *et al.* 2001). This is especially the case nearer towns and cities.

Income Shares and Farm Size: shares of rural non-agricultural income may rise or fall with farm size (falling generally in Latin America, rising in Africa), but tend to rise with household wealth and household income; levels of rural non-agricultural income, on the other hand, rise more often with farm size, and universally with household wealth and income (Reardon *et al.* 2001; Barrett *et al.* 2001). Shares of income from self-employment often rise with farm size (Da Silva and Del Grossi, 2001; Berdegue *et al.* 2001).

Specialization and Wealth: specialization of household activities tends to increase in wealthier zones, when comparisons are made between zones, but to decrease in wealthier households, when comparisons are made between households in the same zone (Reardon *et al.* 2001). Deininger and Olinto (2001) show that Columbia is an exception to this pattern: in Columbia increased specialization and household wealth are positively related.

Effects of Off-farm Income – Food Security: diversification of income sources appears to improve food security (Ruben *et al.* 2001; Reardon and Mercado-Peters, 1993), at least in the short term.

Effect on Farm Investment: the availability of off-farm monies may either increase or decrease on-farm investment and efficiency (Reardon *et al.* 1992; Collier *et al.*, 1986), depending on the agricultural environment.

Effect on Total Income Growth: an initial diversification of income sources tends to improve household income growth over time (Block and Webb, 2001; Canagarajah *et al.* 2001).

Effect on Equality: Results are mixed on the effects of non-agricultural and off-farm income on inequality. In Africa, barriers to entry to well-paid off-farm work appear to lead to an inequality increasing effect of off-farm work. Off-farm wage work, for example, increases inequality while self-employment decreases it (Canagarajah *et al.* 2001). Worldwide, decomposition of Gini coefficients and Shorrock type regressions suggest that results are more mixed (Adams and Alderman, 1992; Lanjouw, 1998; Lazlo, 2001), with the evidence somewhat weighted towards an inequality increasing effect for off-farm work.

1.1.7 Research findings in developed countries using the farm household model

The main findings from the participation literature for developed countries are:

Age: The probability of off-farm employment increases for younger ages, and then decreases for older ages (Sumner, 1982; Benjamin, 1994; Weersink *et al.* 1998). Capital is accumulated over time so that as age increases there is more capital to initiate an enterprise (Evans and Ngau, 1992).

Unearned Income: Almost without exception, it is found that increases in non-agricultural asset income have negative effects on off-farm participation (Bollman, 1991; Huffman and Lange, 1989; Gould and Saupe, 1989; Tokle and Huffman, 1991; Kimhi, 1994). The exception is Huffman (1980).

Experience: Off-farm work experience is significantly correlated with off-farm work participation (Mishra and Goodwin, 1997). A negative relationship exists between farming experience and off-farm work (Mishra and Goodwin, 1997).

Education: Level of education increases participation rates (Gould and Saupe, 1989; Olfert *et al.* 1993; Benjamin, 1994). The effect is stronger for females (Furtan *et al.*1985) as women are more likely to have jobs for which education is important while men usually find employment for which experience but not education is important. Attendance at specialist agricultural courses reduces the participation rate in off-farm labour (Benjamin, 1994; Mishra and Goodwin, 1997)). An increase in spouse's education often decreases participation for the operator (Huffman and Lange, 1989; Tokle and Huffman, 1991; Kimhi, 1994b), although the opposite is sometimes found (Bollman, 1991; Lass and Gempesaw, 1992).

Children: The presence of children in the farm household affects the wife's off-farm work participation rates but has no effect on participation of the husband (Benjamin, 1994). The number of children has a negative effect on participation, which decreases with the child's age (Gould and Saupe, 1989; Olfert *et al.* 1993; Benjamin, 1994; Mishra and Goodwin, 1997; Weersink *et al.* 1998). Increased family size has a positive effect on participation in the off-farm labour market (Woldehanna *et al.* 2000). If both farmer and spouse work, each pre-school child carries a negative effect on off-farm work participation (Lass and Gempeshaw, 1992).

Personality: Farmers committed to farming will gain utility from work on-farm. This carries a negative effect on off-farm employment (Weersink *et al.* 1998).

Farm size: Farm size has a negative effect on farm household participation in the off-farm labour market (Benjamin, 1994). Larger farms have less labour flexibility, reducing the possibility of partaking in off-farm work (Mishra and Goodwin, 1997).

Agricultural Enterprise: Dairy operations have a high labour requirement and therefore have a negative effect on off-farm employment (Gould and Saupe, 1989; Lass and Gempeshaw, 1992; Weersink *et al.* 1998).

Local Market: The quality and the level of development of the local infrastructure will affect the ease of obtaining inputs and marketing outputs. Probability of off-farm work decreases with distance to town (Sumner, 1982; Lass and Gempeshaw, 1992). Greater flexibility will allow optimal hours to be worked and, therefore, increase participation rates (Hearn *et al.* 1996). The regional unemployment rate has a negative effect on off-farm employment (Gould and Saupe, 1989).

Credit: Farm households with higher debt to asset ratios were observed to have a higher participation rate in the off-farm labour market (Mishra and Goodwin, 1997). Weersink, *et al.* (1998) observed that increasing debt to asset ratio or decreasing net farm income reduced participation in off-farm labour markets for the farmer but not the spouse. Additional income from non-agricultural sources helps to obtain basic necessities for those with low farm incomes and improves the standard of living for the farm household in some other cases (Weersink *et al.* 1998).

Risk: Operators, rather than spouses, are found to be more likely to work off-farm to compensate for on-farm income volatility (Mishra and Goodwin, 1997).

Wages: wage elasticity tends to be higher for females than males in both developed and developing countries (Rosenzweig, 1988).

Policy: Direct policy effects on the off-farm decision may come from price supports, grants, extension or credit schemes. Keeney (1999), for example, finds that the more decoupled are direct payments the less is their negative influence on the decision to work off the farm. Indirect policy effects are possible through taxation, education, transport and social policies that vary in their effects on and off the farm.

1.1.8 Main differences between developed and developing country results

In developed countries a negative relationship is clearly evident between off-farm participation and farm size and between off-farm income share and farm size, while in developing countries there is no uniformity in these relationships. In Africa, for example, both relationships appear to be mainly positive, whereas in Latin America, off-farm income share tends to decline with farm size. The African results are likely to be due to higher barriers for well-paid off-farm work in developing countries. Dercon (1998), Dercon and Krishnan (1996) and others argue that there are high capital market imperfections due partly to deficiencies in property rights for collateral, and the accompanying informational assymetries, which prevent entry of smaller farmers to higher paying non-agricultural self-employment in Africa; Carter, *et al.* (1998) argue that there are subtler but strong wealth constraints in South Africa determining possibilities for both waged and self-employed non-agricultural work. There is also a stronger positive relationship between total income and wealth and off-farm income levels and shares in developing countries than in developed countries – again, likely due to higher barriers to entry to better paying non-agricultural work in the former.

A second major difference in the pattern of results relates to household composition variables such as gender of household head or the presence of young children and off-farm participation. In developing countries the presence of young children tends to either have no effect on participation or to increase participation in badly paid off-farm work, with no effect on entry into well-paid work. In developed countries, on the other hand, the effect of having young children on participation is usually negative for the spouse and zero for the operator.

Also, in developed countries the participation rate of females appears to be clearly less than that of males. This difference is not so evident in developing countries outside Asia or in non-Islamic countries. Finally, there is a clear difference in the type of work that males and females enter into – with females tending to enter badly paid self-employment in developing countries (Barrett *et al.* 2001; Woldenhanna and

Oskam, 2001; Ruben and van den Berg, 2001; de Janvry and Sadoulet, 2001; Lanjouw, 1998; Lanjouw 2001) while in developed countries females working off the farm tend on the whole to have relatively well paid jobs, where education is important (Furtan *et al.* 1985).

1.1.9 Policy and research agendas

Simultaneous with this research activity, Reardon *et al.* (2001) call for off-farm and non-farm income issues to be placed at the centre of both the research and the policy agendas for rural development in developing countries. He argues that up to now issues surrounding off-farm income sources have fallen 'in the gap between the institutional walls of governments, research institutions and NGOs'. Because non-agricultural activities are often involved, agricultural departments and research institutions normally ignore off-farm income sources. Because they are small scale, governments and enterprise policy institutions normally ignore them. In Africa, at least, much of the burden of developing off-farm income sources has fallen on NGOs, which themselves often lack the scale to develop enterprises and the resources to co-ordinate them.

Apart from Reardon's demand to stake out a policy and research agenda for rural non-agricultural incomes, the recent findings have led to specific calls on policy makers not to ignore rural off-farm issues in a variety of settings. At times, for example, it may be necessary to remind policy makers that rural education and rural infrastructure are not simply engines to promote agricultural growth but also have important roles to play as generators of non-agricultural household income. Similarly, training, education and credit market barriers to well-paying non-agricultural rural employment may cause even greater aggregate income loss than, say, missing agricultural credit markets or the lack of agricultural extension services. The possibility of lowering barriers to women's participation in growth-generating non-agricultural activities may also be amenable to policy initiatives. Local, regional and national governments will also have a part to play in a wide range of other measures to encourage non-agricultural income generation. These can include taxation changes, licensing arrangements, the direction of public resources to certain areas, the attraction of particular kinds of inward investment, improvements in co-operative marketing, integrated regional planning etc.

Given the present state of knowledge, some of the main tasks of ongoing and future research in the area are as follows:

(i) modelling different kinds of off-farm and non-agricultural labour decisions (e.g. modelling the determinants of participation in different kinds of off-farm work, or modelling decisions of households in different 'regimes' such as, for instance, off-farm participation decisions both for households planning to exit farming and for households planning to increase farm size);

(ii) examining in more detail the effects of barriers to entering higher paying non-agricultural employment on both the poverty alleviation potential of non-agricultural income and on the influence of non-agricultural income on rural growth and rural inequality;

(iii) analysing longer term dynamic strategies of off-farm employment (e.g. can households move from low paying to higher paying non-agricultural employment activities over time, by using money earned from the former to overcome barriers to entering the latter);

(iv) cataloguing and analysing inter– and intra-country differences in the determinants and effects of non-agricultural employment, income levels and income shares (e.g. there appears to be a stronger link between off-farm income and household income inequality in Africa than in Latin America – (Reardon *et al.* 2001) – or there may be different effects of increases in off-farm income on agricultural investment strategies, depending on the agro-ecological environment (Pascual and Barbier, 2001);

(v) developing more sophisticated models of farm household decision making which allow for the interdependence within households of certain family members' decisions or, on the other hand, for the interdependence of decisions among different households within a village or rural community (Taylor and Adelman, 2002).

1.2 The collective model: introduction

In this section of the literature review chapter some of the main features of the Collective utility model are outlined. Most of the results of tests have been for OECD countries and an exhaustive list is not provided. But the most important aspects of the literature are treated. The Collective model has not been part of the farm household literature until very recently, but the likelihood is that this will change.

1.2.1 Collective household model

Two implications of the Unitary Utility model for households are income pooling[5] and Slutsky symmetry. Pooling of income by household members means that the sources of household income should not affect allocation decisions, including labour supply. Slutsky symmetry is a non-intuitive implication of the Unitary model which states that Hicksian cross-price effects should be symmetric. Research by many authors (Altonji *et al.* 1989; Cai, 1989; Schultz, 1990; Thomas, 1990; Bourguignon *et al.* 1993; Phipps and Burton 1992; Lundberg *et al.*, 1977) has cast doubt on the income pooling hypothesis, using mainly unearned income in their tests, to avoid problems of endogeneity. The Slutsky restrictions, which impose the symmetry of cross wage effects on the compensated labour supply of each household member, have also been rejected (by, among others, Keeley *et al.* 1978, Ashworth and Ulph, 1981; Kooreman and Kapteyn, 1986; Browning and Meghir, 1991 and Zhang and Fong, 2001).

Such rejections led Chiappori (1988) and Apps and Rees (1988) to formulate what has become known as the Collective model. This is the main alternative in the empirical literature to the Unitary model. The aim of Chapter 3 is to decide which of these models is most appropriate for rural Peru.

In the Collective model, each household is assumed to have (normally) two partners. Each partner is assumed to have their own utility function, U, which is a function of both own leisure and consumption and partner's leisure and consumption. The version of the model based on these kinds of utility functions is usually called the 'general' model. It is important to note that this version of the household utility function implies that externalities exist between partners. The operator's utility, for example, can be affected by what the spouse consumes and by how much leisure she has, and vice versa. It is likely that externalities do exist between household members, if only because they live in such close proximity to each other, so this would seem an

[5] .....though Browning, Chiappori and Lechine, 2004, argue that the Slutsky matrix conditions tested in Chapter 3 are in fact the signature of a Collective model; in their definition, non-pooling of income is possible in a Unitary context and income pooling in a Collective one

intuitively plausible formulation for household utility. It is not necessarily a purely altruistic utility function, although (more restrictive) altruistic utility functions can be used in the 'caring' version of the Collective model, which will be discussed below.

Let h be hours worked, C be consumption, w be wages, y be unearned income, z be variables that affect preferences, s be distribution factors (variables that affect intra-household bargaining strength, but not preferences) and the i subscript, or subscripts 1 and 2, represent the two partners. For now, all households are assumed to have two wage earning members. Household production is assumed to be zero – all time not working for money is leisure time.

The identifying assumption of all versions of the Collective Model is that household decisions are Pareto-Efficient. The Pareto Efficient outcome is justified by the likelihood that if household decisions are the outcome of a co-operative game, or of a repeated non-cooperative game, Pareto Efficiency will result. The Collective Model is usually agnostic about the type of bargaining that goes on within households (Manser and Brown, 1980; McElroy and Horney, 1981), though occasional reference is made to differing bargaining models when the data allows – Chiappori *et al.* 2001; Rapoport *et al.* 2003.

Given a Pareto-Efficient outcome, for any given (w_1, w_2, y, s, z), there exists a weighting factor $\mu(w_1, w_2, y, s, z)$ between 0 and 1, such that (h_i, C_i) solves the following:

$$\max_{(h^1,h^2,C^1,C^2)} \mu U^1 + (1-\mu)U^2$$

s.t.

$$w_1 h_1 + w_2 h_2 + y \geq C_1 + C_2$$

and $$0 \leq h_i \leq 1, i = 1,2 \qquad (1-14)$$

where the weighting function is a continuous and differentiable function of the exogenous variables w, z and s. All the exogenous variables (except s) affect both the position *of* the Pareto frontier and the position *on* it. The importance of the distribution factor – s – is that it only affects the final position *on* the Pareto frontier, and not the position of the frontier itself. This is important for testing for the model.

Chiappori and Ekeland (2001) show that the following equality of ratios is both a necessary and sufficient condition of the Collective Model:

$$\frac{\partial h_1 / \partial s_k}{\partial h_1 / \partial s_1} = \frac{\partial h_2 / \partial s_k}{\partial h_2 / \partial s_1}, \forall k = 2,....,L \qquad (1-15)$$

where there are L distribution factors, each subscripted by k. The equality can be tested for using survey data, as long as there is information on more than one distribution factor.[6] The test is not valid if s also affects preferences (e.g. if s includes variables such as education or age) since it is only *along* the frontier that inter-personal marginal rates of substitution are equalized and the equality of the ratios holds. External factors (such as spousal wealth on separation – Mendoza (2002); type of divorce legislation or the local sex ratio– Chaippori, *et al.* (2001); bridal wealth – Thomas (1997); single parent allowances – Rubalcava and Thomas; pension recipients– Duffo; and local variation in alimony rights– Rangel, 2003) may be usable as distribution factors.

For the general version of the model, own or spousal wages are not useable as distribution factors, nor is household non-labour income, since these variables directly affect the position of the frontier. This can be best seen when examining Marshallian labour supply functions: $h_i = H_i(w_1, w_2, y, z, \mu(w_1, w_2, y, z, s)), i = 1,2.$ (1 – 16)

Differentiating by s gives us: $\frac{\partial h_i}{\partial s_j} = \frac{\partial H_i}{\partial \mu}\frac{\partial \mu}{\partial s_j},$ (1 – 17)

which in turn implies the Chiappori and Ekeland equality (1 – 15). The important point is that the equality of ratios is independent of i. This would not be the case were one to differentiate Marshallian labour supply with regard to w_1 or w_2 or y or z, or – if the data were available and one could consider doing so – by y_1 or y_2 instead of y.

For unique identification of μ to be possible one must assume egoistic or 'caring' preferences, where each persons' utility is a function of own consumption and leisure only, or of own consumption and leisure, plus partner's utility – but not of the partner's actual leisure and consumption.

In the case of the caring version of the Collective model, the weight given to partner 1's utility – μ – takes on a new, and convenient, meaning, notation usually changes

[6] A similar test can be carried out for consumption goods instead of labour supplies (Bourguignon *et al.* 1993). The equality of marginal rates of substitution between individuals, which is the basis on which these tests are built, is a standard implication of Pareto Efficiency. A further set of tests for the Collective Model in its general form, with or without distribution factors, is developed in Browning and Chiappori (1998).

and the focus moves from utility to income. Pareto efficiency with caring (or egoistic) preferences implies:

$$\max_{\{h^i, C^i\}} U^i(1-h^i, C^i, z)$$

s.t.

$$w_i h^i + \phi^i \geq C^i$$

$$0 \leq h^i \leq 1 \qquad (1-17)$$

where $\phi^1(w_1, w_2, y, z, s) = \phi$ and $\phi^2(w_1, w_2, y, z, s) = y - \phi$. $\qquad (1-18)$

$\phi^1(w_1, w_2, y, z, s)$ is the Lagrangian multiplier on the constraint imposed by Pareto Efficiency that person 2's utility must be above a fixed level (determined by the exogenous variables), so that person 1 can only maximize their utility without diminishing that of person 2. From the formulation above, ϕ can then be interpreted as a sharing rule. Non-labour income is (analytically) first divided between the couple according to the shares given by the rule. The couple then choose labour supplies and consumption to maximize their individual utility functions subject to their first step budget constraint, given by the sharing rule. Chiappori *et al.* (2001) prove that several empirically testable, results follow.

Let $h^1 = H^1(w_1, \phi(w_1, w_2, y, s, z), z)$ and $h^2 = H^2(w_2, y - \phi(w_1, w_2, y, s, z), z)$. Given maximization of the utility function, and interior solutions, these are the Marshallian labour supply functions for the couple using the caring model. Changes in the spouse's wage, for example, have an effect on the operator's labour supply only through the sharing rule (as do changes in y or s): in other words, changes in w_2, y and s only have an income effect on partner 1's labour supply. No substitution effects exist, except from changes in own wages. Formulated in this way, the previous test for the Collective Model (1 – 15) could simply be repeated, replacing a second distribution factor by non-labour income, and so only one distribution factor is needed. However, in a caring utility framework one can go further.

Specifically, the following set of consequences result: let

$$A = \frac{h^1_{w_2}}{h^1_y}, B = \frac{h^2_{w_1}}{h^2_y}, C_1 = \frac{h^1_{s_1}}{h^1_y}, D_1 = \frac{h^2_{s_1}}{h^2_y}, \qquad (1-19)$$

where the 1 subscript refers to the distribution factor. If $h_y^1 * h_y^2 \neq 0$ and if a distribution factor exists such that $C \neq D$, then Chiappori *et al.* (2001) list eight conditions, analogous to the Slutsky restrictions of the unitary model, which are necessary for any pair of unrestricted labour supplies to be the solution of the sharing rule. Three of these conditions are presented here simply to give a flavour of what is involved. Only one distribution factor is assumed (so dispensing with the subscript):

(a) $\frac{\partial}{\partial s}\left(\frac{D}{D-C}\right) = \frac{\partial}{\partial y}\left(\frac{CD}{D-C}\right)$

b) $\frac{\partial}{\partial w_1}\left(\frac{D}{D-C}\right) = \frac{\partial}{\partial y}\left(\frac{CD}{D-C}\right)$

c) $\frac{\partial}{\partial w_1}\left(\frac{CD}{D-C}\right) = \frac{\partial}{\partial s}\left(\frac{BC}{D-C}\right).$ (1 – 20)

If the eight conditions hold, then the sharing rule is defined up to an additive constant, depending on z. Its partial derivatives (again, with one distribution factor) are:

$$\phi_y = \frac{D}{D-C}; \phi_s = \frac{CD}{D-C}; \phi_{w_1} = \frac{BC}{D-C}; \phi_{w_2} = \frac{AD}{D-C}. \quad (1-21)$$

Even if there are no distribution factors, Chiappori (1988, 1992) shows that the sharing rule can still be recovered (again, up to an additive constant), but in this case by using functions of the derivatives of A and B (hence functions of the second derivatives of labour supplies). The existence of at least one distribution factor thus has the effect of making recovery of the sharing rule more 'robust', since – as we see above – A, B, C and D themselves can be used to recover the sharing rule (up to an additive constant).[7] Likewise, the eight testable predictions concerning the adjusted Slutsky matrix only require first derivatives of (multiples of) A, B, C and D, and not the rather complex second derivatives required if there are no distribution factors.[8]

---

[7] If the focus is on consumption rather than labour supply, then – without distribution factors – it is possible to recover the sharing rule in a similarly more 'robust' way – again, up to an additive constant – as long as there is consumption information on an 'assignable' or 'exclusive' good, as well as expenditure information on other goods (Browning *et al.* 1994, with particularly clear expositions in Deaton, 1998, pp226-228 and in Apps and Rees, 1995). But in (almost) no case are the shares themselves recoverable, because individual consumptions are almost always lacking in the data. The one exception in the literature so far is Couprie, (2003), who uses panel data on movements into and out of marriage to identify actual intra-household consumption shares in the UK (consumption by a person when single is used to help identify consumption of the same person when married).

[8] Tests, without distribution factors, based on these third derivatives were first proposed in the seminal paper by Chiappori, 1988.

These are remarkable results, and there has been an explosion in the recent literature in recovering sharing rules (up to a constant) in a variety of settings (France, by Browning, *et al.* (1994); Canada, by Fortin and Lacroix (1997); Italy, by Chiuri (1999); Australia, by Blacklow and Ray (2003); India, by Lancaster, *et al.* (2003); Spain, by Zamora (2003); the UK, by Couprie (2003); Switzerland, by Gerfin and Wanzenreid (2003); Holland, by Vermuelen (2005)). However, a variety of problems still exist in estimating Collective Models, which is not surprising, given that this is an extremely new literature.

The problems that remain include a) the case of non-participation, which renders the two step optimization story of the caring version inapplicable, since separability no longer holds;[9] b) non-linearity of returns (e.g. work that is in self-employment, or has non-linearities induced by transaction costs and non-separabilities, which will be discussed Chapter 5 below) and c) the consumption of public goods or externalities in private consumption (e.g. children, leisure time), which also mean that the two step analogy is inapplicable.

Each of these difficulties is related to the other, and each has received some treatment in the literature, especially the first two, since they also cause difficulties when estimating labour supply using the Unitary Model. We will look at each in turn.

(a) *non-participation:* in the Unitary Model there is a particular problem in dealing with non-participation of one or more partners. If the spouse participates, we have seen that the operator's reservation wage is conditional on the spouse's wage, as well as all other exogenous variables (bar the operator's own wage). If the spouse does not already participate, exogenously as it were, then the operator's participation decision can only be made simultaneously with the spouse's decision. This leads to a practical problem, mentioned in section 1.1.3 with the commonly used bivariate probit formulations to model the couple's participation decisions. With regard to the Collective Model, there is the added issue of the sharing rule. Blundell *et al.* (2002) show that in the Collective Model Pareto Efficiency implies that BOTH partners must be indifferent to a particular partner working or not, if the

[9] Furthermore, in attempting to model the participation decision itself it must be remembered that the reservation wage is not the same concept in the Collective Model as it is in the Unitary Model, since for each partner (if the other partner is working for a given wage) there is, in effect, a reservation frontier, rather than a single reservation wage, as different possible sharing rules mean that the individual's participation decision depends on a series of combinations of reservation wages/sharing rules.

offered wage for that partner is at their reservation wage (e.g. if the spouse is indifferent between working off-farm or not, then the operator must also be indifferent between the spouse working off-farm or not). Furthermore, the wage offered to the non-participating partner is an element in the participation decision of the other partner, since the offered wage, even if the partner does not participate, is likely to affect the balance of power within the family. Blundell *et al.* (2002) outline a number of restrictions, based largely on the above two points, that can be tested to compare the applicability of the Unitary and Collective models. They also recover the sharing rule. Donni (2001) also recovers the sharing rule, and generalizes the results of Blundell *et al.* (2002) to include non-participation by one or other partner (Blundell *et al.* (2002) are only concerned by female labour supply in the face of a discrete choice by the male). The result of this work is that Chiappori's original identification and testing results are extended to the participation frontier and that additional tests can also be generated by the boundary conditions themselves.

(b) *non-linearities or lack of separability*: in many developing (and developed) countries market imperfections of various types can lead to lack of separability between production and consumption decisions. This issue is the explicit focus of Chapter 5, and will be discussed there.

(c) *consumption of public goods or externalities in private consumption:* the first of these problems has been treated in a paper by Chiappori, Blundell *et al.* (2002), and by work by Donni (2004) and Couprie (2003). The approach by Chiappori *et al.* has been to modify the two-step process, so that in the first step joint expenditure on the public good(s) and the sharing rule for residual income is decided on; in the second step, as usual, individual leisure and private consumption are decided. The double decision of the first step leads to testable implications. Recoverability of the sharing rule is conditional on having time use data, and on having information on public expenditure (e.g. expenditure on children), and – if there are no distribution factors – on quite a strong separability assumption (that the consumption / leisure trade off is not affected by consumption of the public good). The presence of distribution factors renders the separability assumption unnecessary – and so is another reason that one might seek to use distribution factors, if data allows. Couprie recovers the sharing rule (and, for the first time in the literature, the actual shares) using change of marital status in a household panel data set to identify individual leisure demands. Donni, on the other hand, uses conditional demand functions (demand conditional on exclusive goods) to recover both private and public demands.

Finally, if there are externalities in private consumption, the Collective Model can be tested for using the Slutsky matrix approach used in Chapter 3, or – if there are more than one distribution factors (or assignable goods)– by testing for the equality of distribution factor derivatives noted at the beginning of this section. If there are consumption externalities, however, an income sharing rule does not make sense (we are back to the general form of the Collective Model, with weights on utilities, not money), and a sharing rule is simply not recoverable.

1.2.2 The collective model and the farm household labour supply literature

The only example in the literature at the time of writing of a collective model being used for actual farm household labour supply estimations[10] is a reduced form model by Mendoza (2002), who estimates off-farm labour participation in the Philippines. However, a number of tests for the Collective model have been carried out. The identifying implication of the collective model for testing is that household allocations must be Pareto efficient (because they are outcomes of a co-operative game). Thomas and Chen (1994) find against the unitary model using Taiwanese data, but cannot find against Pareto efficiency (using tests based on Equation 1 – 14), and so support the Collective model, as have a number of studies in developed countries (Browning, Chiappori *et al.* 1994; Chiuri, 1999; Fortin and Lacroix, 1997). However, Udry (1996) finds against Pareto efficiency in household plot allocations in Burkina Faso (a finding later reversed, for neighbouring villages, by Akresh, 1999) and Duffo and Udry (2000) also find against it. Fortin and Lacroix (1997) find against Pareto efficiency in the labour allocation decisions of Canadian parents with young children, but not for parents without children. Partially in response to these findings, Basu (2004) argues that if dynamics are introduced into the model (by allowing present actions to have a feedback effect of on future bargaining strength) then sub-optimal

[10] Appendix 1 contains a page or two on how the Collective model might be extended to relate to the farm household scenario.

outcomes are possible in the collective model. A series of estimations have been used to recover the sharing rule following on his work (Blacklow and Ray, 2003).

1.3 Conclusion

This chapter has introduced the basic framework that is used by economists to analyze a variety of decisions by farm households. This is the farm household model, a development of consumer theory that lies squarely within the neo-classical tradition. It went on to discuss a number of estimations of labour supply and off-farm participation in developing countries that have been carried out within the general framework of the farm household model, as well as more descriptive models of income diversification and specialisation. In the last part of the chapter the Collective utility model was introduced, with the focus also on labour supply. This is another extension of consumer theory used to deal with aggregation of preferences within households in general, not specifically farm households.

In Chapter 3, the semi-reduced-form[11] utility models (Unitary and Collective) that may underpin the farm household model will be compared for farm households in Peru. In the concluding Chapter (7.1 and 7.3) the importance of utility model issues for further research and for policy are discussed. In all the estimations in the other chapters, the farm household model is the guiding model.

[11] Semi-reduced forms because, on the one hand, the type of bargaining, or ways in which preferences or aggregated that lie behind either model are not spelled out but, on the other, enough is knowable about the welfare of final outcomes for one to say that all structure has not been 'reduced away'.

2

Background and Data

The aims of this chapter are to explore the dataset that is used in the chapters that follow and to provide some background on the area of study – rural Peru.

2.1 The dataset and rural Peru

The Living Standards Measurement Survey (LSMS) of 1994 provides detailed information on household income and expenditure in urban and rural Peru. The 1994 survey covers the whole national territory. The information includes basic personal information on all household members, as well as information on their dwelling places. It includes personal information on education, health, hours and place of work, income, self-employment information (on the main non-agricultural family business, if there is one), detailed household expenditure information, information on migration, on household savings and credit as well as information on agricultural production, inputs, outputs and 'self-consumption'. In most of the rural regions community level questionnaires were also completed through interview with local leaders. These provide information on local wage rates, local infrastructure and employment activities and, for many communities, local prices.

The World Bank has made this data available on the web and it has been used, along with data from other Peruvian LSMS surveys in 1985, 1990, 1991 and 1997 (the latter not publicly available), in studies concerning poverty alleviation policies (Schady, 2000), household adjustment to macro-economic shocks (Schady, 2002), poverty and inequality (Robles, 1997), labour supply in rural households (Jacoby (1993) using 1985 data – a seminal work, mentioned in the previous section, in the literature of non-separability of household labour), labour supply in rural households again (Newman and Gertler (1994) using 1985 data), the determination of off-farm income shares (Escobal (2001) using 1997 data) and, most recently, the returns to education (Lazlo (2002) using the 1991 data), farm household labour supply (Lazlo, 2002) and investigating separability (Vakis *et al.* (2004) using 1997 data).

The full 1994 data set has information on 3958 adults in farm households (over 14 years old). The quality of LSMS data is generally held in high regard (Deaton, 1999), but in certain cases in constructing the dataset it has been found that implausible values need to be adjusted for (e.g. the occasional person who works more than 24 hours a day), and it is not certain than even plausible values do not in some cases need correction. Nevertheless, the data from the surveys is very comprehensive in its household sections, and reasonably so in the agricultural sections. Only those rural areas for which community level questionnaires were given are included in this study, so that information on local infrastructure and economy is available for all households.

Before exploring some of the data from the LSMS surveys background information on agriculture in Peru is presented.

2.1.1 A brief description of agriculture in Peru

In 2000, the population of Peru was around 27m. About 36% of Peru's population live in the sierra (the Andean mountainous region), with about 12% in the Amazonian rainforest, and the rest in the coastal region and in Lima and the surrounding areas. Roughly 45% of the population are Amerindian, which is high for Latin America (only Bolivia and Guatemala are comparable). Mestizos (persons of mixed Amerindian-European ancestry) constitute 37% of the population. Persons of pure European ancestry constitute about 15%, while about 3% are of mixed nationality, mainly East Asian and Afro-Peruvian. The main languages are Spanish and Quechua (the principal indigenous language). Some of the indigenous groups have been assimilated into mestizo culture but, in the East Andes especially, traditional customs are more likely to be adhered to.

In 1994, the year of this study, income inequality, as measured by the Gini coefficient, was .449 for the country as a whole, and .35 in urban areas, indicating an extremely high degree of inequality in rural areas (Deininger and Squire Inequality Dataset – World Bank). Leading up to 1994, general economic expansion in the mid 1980s had been followed by collapse and default in the late 1980s and an upsurge in unemployment and inflation, and in guerrilla activity. By the time of the LSMS survey in 1994, however, most guerrilla activity had ceased, and the economy was growing fast. From the 1960s to the 1990s the long process of progressive urbanization had also continued. Some bare facts are illustrated in Tables 2-1 and 2-2 below.

Table 2-1 : GDP and inflation growth in 1990s

	GDP Growth
1992	-0.4%
1993	4.8%
1994	12.8%
1995	8.6%
1996	2.5%
1997	6.7%
1998	-0.5%
1999	0.9%
	Inflation
1990	7669%
1994	15.4%

Source: INEI (Peruvian Statistical Agency)

Table 2-2: Urban / rural population shares over time

	Rural	Urban
1940	64.6%	35.4%
1961	52.6%	47.4%
1972	40.5%	59.5%
1981	34.8%	65.2%
1993	29.9%	70.1%
2002	27.8%	72.2%

Source: INEI (Peruvian Statistical Agency)

Figure 2-1 below maps the urban/rural population share across the country.

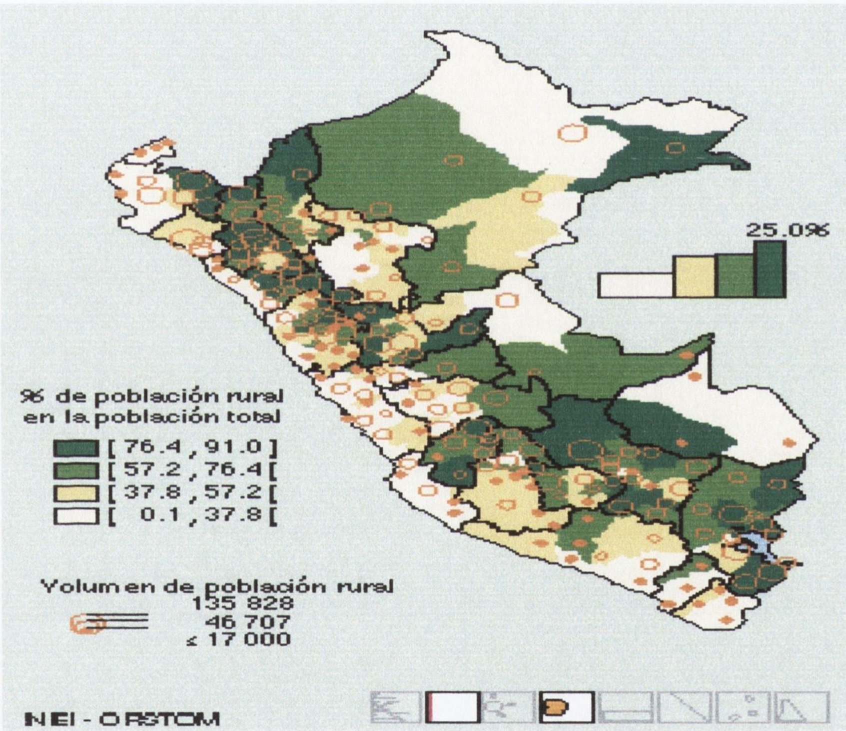


Figure 2-1: Urban/Rural Population Shares, 1993 (source: INEI internet site)

The main concentration of the rural population in 1994 was on small farms in the rural highlands, the majority operated by Quechua speaking households (and other households where Spanish was not the primary language) with more commercial, generally profitable farming on the coast and larger scale but low-return farming in the forested areas of eastern Peru. Both coastal and forest areas are inhabited mainly by Spanish speakers.

The process of land reform, initiated and then abandoned after an uprising in La Convencion in 1962, was re-inaugurated and followed through on a national basis by the military Velasco regime (1969-1975), based on the recommendations of the CIDA (Inter-American Committee for Agricultural Development) report on land reform in seven Latin American countries (Peru's was published in 1966). About half of all agricultural land was expropriated, with about 33% of agricultural households benefiting from reforms. Later, in the 1970s, excluded households (members of mainly indigenous peasant communities) also gained rights to expropriated lands, after many local uprisings and peasant occupations of reform land. These groups had been granted rights to shares of – largely non-existent – profits in the main reform package of the late 1960s, but not rights to land.

For the first twenty years after reform much of the expropriated land was given to cooperative farming organisations – about half –, most of which were in the coastal regions. Most of the more profitable haciendas had been along the coast, where pre-reform peasants had worked for wages, rather than under the more feudal arrangements of the pre-reform highland peasants. All forms of collective farming, however, gradually dissolved during the 1980s, leaving a mainly individualised sector operating by 1994. Almost 80% of total land was owned by men, and 20% by women (Census, 1994). Land sales were few, and land rental almost non-existent, until further reform in 1995 by the Fujimori government allowing foreign ownership, large scale farming and, in particular, instituting a cheap registration programme. The initial reforms had led to very little if any increase in productivity (van der Ploeg, 1990).

As part of a wider privatisation policy, the Fujimori government also liberalised all prices of inputs and outputs from 1992 on, reduced farm extension programmes, eliminated state ownership of the main input industries (chemicals and fertilizer) and closed down the main state-backed agricultural bank, the Banco Agrario, which between 1985 and 1990 had lent nearly $3bn to the agricultural sector at negative real interest rates, replacing it with a battery of smaller, market driven local credit agencies (Cajas Rurales de Ahorry y Credito, many of which had run into trouble by the late 1990s).

The household economy in the rural Andes has been the subject of much anthropological research (Bourque and Warren, 1981; Deere and Leon, 1982; Deere, 1986; Nash, 1986; Fernandez, 1988; Flora and Santos, 1986; Alberti, 1988; Jacoby

1991; Hamilton *et al.* 2001). Most of the work has been in the form of case studies, so generalisations are quite difficult. With regard to farm work, women tend to look after livestock and men tend to look after both livestock and crops. Planting decisions are often made together in the sierra (Jacoby, 1991), but men tend to have more overall control of farm decision-making the more market-oriented is the farm (Nash, 1986; Alberti, 1988). In the sierra, women tend to be self-employed – in crafts, weaving, selling – and work in other farms less than men.

Traditionally, in most of the indigenous cultures, a two-headed household is desirable, where most income is pooled and most decisions joint. In these communities lands tends to be divided equally among children (Hamilton *et al.* 2001). This is not the case within mestizo households or others strongly influenced by Hispanic culture (Deere and Leon, 1982), where primogeniture holds sway (even if Hispanic culture is both matrilineal and patrilineal). In these rural households also, separate spheres are more apparent for men and women's work, where the men tend to take care of most of the decisions that relate to money and the women look after the home and family.

In the sierra, households tend to be divided into: los pobres (the poor), los medios and los ricos (van der Ploeg, 1990). The poor households tend to work off farm and to have a small amount of land, less than a hectare. The 'medios' are self-sufficient in land, while the 'ricos' own livestock and sometimes trucks, and tend to hire in labour. For 1994, throughout Peru as a whole, the off-farm and non-agricultural work of operators takes place in the sectors shown in Table 2-3.

Table 2-3: Sectors people work in off the farm

Operators Leaving Farm for Work		Operators Diversifying on the Farm	
Sectors		Activities	
Other Farm work	166,981	Crafts	86,962
Collecting Fruit/Hunting	10,304	Trade	42,259
Wood Extraction	13,153	Derived Products	37,111
Fishing	17,712	Machine Goods	15,877
Mining etc.	18,490	Renting Machines	2,004
Manufacturing	4,795	Food and Drink	1,248
Construction	45,146	Coal	384
Trade	68,235	Carpentry	433
Catering	2,419	Regional Products	829
Transport	16,555	Other	90,487
Teaching	12,195	Not Specified	10,610
Domestic Service	33,935		
Not Specified	27,954		
Total:	451,761	Total:	283,345

Source: 1994 Agricultural Census

The principal crops of Peruvian farmers in 1994, and the farm size and tenure distribution are shown in Table 2-4.

Table 2-4: Farm Size, Tenure and Main Crops

Total agricultural area (hectares)	35,637,808		
Number of operators	1,742,267		
Number of 'parcelas':	5,718,079		
Farm size (hectares)	**% of farms**	**Main crops**	**% of area used for non-permanent crops**
< 3	53	Cereals (maize is the largest)	42.5%
3 – <10	33	Fruit	8.9%
10 – 50	10	Green vegetables	3.9%
≥ 50	1.9	Beans	6.9%
Tenure	**% of 'parcelas'**	Tubors (especially potatoes - papas)	25.2%
Owned	71.6	Industrial crops	10.7%
Rented	2.4	Others	1.9%
Common use	23.2	Total hectarage on non-permanent crops	2,113,619
Other	2.8	Total hectarage on permanent crops	463,156 (Mainly Coffee - about 100,000)

Source: 1994 Agricultural Census

The farm size distribution is quite unequal and there is a large amount of regional variation in land use. Using data from the 1994 Living Standards Measurement Survey, Table 2-5 outlines the main crop varieties, for the three large regions – the sierra, the cost and the jungle or rainforest.

Table 2-5: Crops grown in the three main Peruvian regions

Region	Crop	Average Regional Yield (kg p/h)	% of cropped area
Sierra	Maize	125.6	37.9
	Papa	389.0	20.7
	Wheat	131.0	19.5
	Cebada Grano	73.7	14.6
	Palto	7.5	2.7
	Total Sampled Crop Area (h)		11,075.4
	No. of Farms in Sample		667
	Median Farm Size		1 hectare
Coast	Maize	507.0	56.0
	Maize (Chala)	351.6	13.4
	Zapallo	192.5	10.2
	Rice	7,741.9	5.1
	Camote	1,104.6	4.7
	Esparrago	1,028.8	3.2
	Alfalfa	5,849.0	2.5
	Total Sampled Crop Area (h)		2,199.7
	No. of Farms in Sample		193
	Median Farm Size		2 hectares
Rainforest	Yuca	147.3	52.9
	Cacao	9.9	16.4
	Platano	749.1	9.8
	Mani Fruta	8.8	5.0
	Rice	2,787.3	3.1
	Pina	5.8	2.0
	Calabaza	14.0	2.0
	Maize (Duro)	645.2	1.6
	Naranjo	743.0	1.5
	Total Sampled Crop Area (h)		10,278.3
	No. of Farms in Sample		398
	Median Farm Size		5 hectares

2.1.2 The LSMS data

In the tables that follow some of the information on household level and individual level variables in 1994 is summarized. These tables are constructed from the main sample base used for the participation and hours estimations of the rest of this section. The sample sizes used are considerably lower than mentioned in the introduction to this chapter. Households which do less than 200 hours per annum in total of on-farm work have been excluded because it was felt that, for households who do very little farm work, no farm level variable could be exogenous. Households producing no crops have also been dropped. In general, also, households where information was felt to be untrustworthy (e.g. anyone working more than 6,000 hours per annum) have been dropped.

The information in the following four tables is taken from the final samples used for most of the estimations in this thesis:

Table 2-6: Main summary statistics

	All households		Households who do more than 100 hours off-farm	Households working less than 100 hours off-farm
	Mean	Std.Dev.	Mean	Mean
Total expenditure (soles)	4734.16	3516.21	4,919.08	4,320.99
Total expenditure per capita (soles)	1195.97	963.90	1,147.97	1,303.23
Share of family farm income	0.60	0.39	0.43	0.97
Share of farm wages	0.12	0.25	0.17	0.02
Share of manufacturing self-employment	0.07	0.21	0.10	0.00
Share of manufacturing wages	0.06	0.19	0.09	0.00
Share of services self-employment	0.10	0.24	0.14	0.01
Share of services wages	0.05	0.17	0.07	0.00
Off-farm hours	1337.20	1721.38	1,934.02	3.69
On-farm hours	3288.82	2469.45	2,931.29	4,087.66
Age	38.31	12.11	36.20	43.04
Gender (1=male)	0.51	0.20	0.52	0.48
Years of education	6.90	3.22	7.32	5.94
On-farm experience	14.80	12.63	12.77	19.35
Farm tenure (higher means less secure)	1.81	1.58	1.96	1.47
Farm capital	758.41	1417.49	622.74	1,061.54
Farm size	4.30	8.32	4.22	4.49
Number children (under 15 years)	2.42	1.87	2.60	2.03
Number of adult family members (15-65)	2.94	1.51	3.10	2.57
Number of people over 65 in the house	0.21	0.51	0.17	0.31
Unearned income (irregular)	245.63	667.94	201.95	343.21
Coastal region	0.16	0.37	0.16	0.17
Rainforest region	0.27	0.45	0.30	0.21
Spanish speaking	0.59	0.49	0.58	0.60
Crop diversification index	2.17	1.09	2.14	2.23
Proportion of crops sold	0.50	0.37	0.48	0.55
Existence of informal labour exchange in community	0.60	0.44	0.60	0.58
N	980		677	303

Table 2-7: Mean hourly returns for family members in four different off-farm categories

	Male main person	Main female	Non-main male	Non-main female
Farm wage work	0.99 (185)*	0.97 (60)	0.88 (100)	0.59 (18)
Non-farm wage work	1.51 (133)	1.18 (31)	1.15 (84)	1.67 (54)
Self-employed (no paid staff)	2.13 (77)	0.14 (68)	2.41 (8)	0.02 (22)
Self-employed (paid staff)	1.29 (108)	0.51 (66)	1.07 (15)	0.25 (23)

*Number of people in each category in parentheses, and top five returns percentiles omitted

Monetary values in the above Tables are in 1994 soles. 2.2 soles bought one U.S. dollar in June 1994. The income share means and standard deviations are unweighted by household earnings.

It can be seen from Table 2-6 that the average 'share' of off-farm agricultural earnings is about 40% of total expenditure, rising to close to 60% for the households who do work off-farm. From an examination of some of the other figures in the tables it appears that those households that tend to work off-farm are households where farming is less commercial than average (a lower proportion of crops are sold: 48% compared to 55%). They are households also where farm size, farm capital and the quantity of hired in farm labour are all lower than the average. However, they are also households where household total expenditure is slightly higher than average but per capita expenditure lower than average (they tend to have more working age adults and more children, but less people over 65).

Longer experience working on the farm also appears to be clearly associated with the likelihood of not working off-farm.

The individual level table (Table 2-7) suggests that operators are more likely to work off the farm, while spouse and other adult males and females are less likely to do so. The returns information in Table 2-7 is also interesting. Working off the farm for farm wages appears to be the least rewarding activity generally for operators (1.15 soles per hour for operators, compared to 1,51 soles in non-agricultural wage work, 2.12 soles in sole trader self-employment and 1.29 in larger scale self-employment). The other sectors are fairly evenly rewarded (at least when samples are reasonably large). For spouses and adult children working for agricultural wages pays less than working for non-agricultural wages, but generally more than self-employment returns, at least for females, which appear in some cases to be suspiciously low (perhaps because of some miscalculation of hours worked by survey respondents).

Table 2-8: Summary statistics by expenditure per capita quintiles

	Mean				
	Q1	Q2	Q3	Q4	Q5
Total expenditure (soles)	2,244.14	3,192.56	4,223.13	5,458.01	8,552.96
Total expenditure per capita (soles)	438.20	714.79	952.09	1,305.02	2,569.76
Share of family farm income	0.60	0.58	0.57	0.59	0.64
Share of farm wages	0.18	0.15	0.14	0.09	0.05
Share of manufacturing self-employment	0.07	0.07	0.07	0.08	0.06
Share of manufacturing wages	0.06	0.08	0.05	0.07	0.05
Share of services self-employment	0.05	0.08	0.11	0.14	0.12
Share of services wages	0.03	0.04	0.06	0.04	0.07
Off-farm hours	1,141.39	1,207.38	1,538.61	1,409.52	1,389.08
On-farm hours	4,007.74	3,093.09	3,530.88	3,158.55	2,653.83
Age	36.59	38.44	37.25	37.41	41.87
Gender (1=male)	0.51	0.50	0.51	0.51	0.52
Years of education	5.68	6.05	6.64	7.38	8.73
On-farm experience	16.34	15.23	13.50	13.67	15.26
Farm tenure (higher means less secure)	1.82	1.85	1.79	1.99	1.60
Farm capital	671.69	580.50	597.90	679.37	1,262.60
Farm size	3.01	3.60	4.02	4.36	6.54
Number children (under 15 years)	3.30	2.70	2.31	2.30	1.49
Number of adult family members (15-65)	3.34	2.88	3.08	2.89	2.50

	Mean				
	Q1	Q2	Q3	Q4	Q5
Number of people over 65 in the house	0.20	0.27	0.21	0.17	0.22
Unearned income (irregular)	199.13	208.56	289.33	225.20	305.91
Coastal region	0.06	0.09	0.17	0.19	0.30
Rainforest region	0.25	0.28	0.29	0.31	0.23
Spanish speaking	0.38	0.55	0.58	0.67	0.76
Crop diversification index	2.36	2.27	2.11	2.11	2.01
Proportion of crops sold	0.36	0.43	0.51	0.58	0.63
Existence of informal labour exchange in community	0.67	0.71	0.61	0.60	0.39
N	196				

In Table 2-8 households are summarized by per capita expenditure quintile, which is often taken in the literature as an indicator of permanent income. The share of off-farm income appears a little higher for low expenditure quintiles (it is over 40% for the four lowest quintiles, falling to 36% for the highest quintile). Off-farm hours are fairly constant for the two lowest quintiles, and higher for quintiles 3, 4 and 5, being especially high for quintile 3. When returns are calculated from the hours and shares information, one finds that they are a little over a sole an hour for quintiles 3 and 4, climbing to more than 1.50 an hour for quintile 5 and falling to less than a sole an hour for quintile 2 and less again for quintile 1. Farm capital is fairly even through the first 4 quintiles, but rises for quintile 5. Farm size rises through the quintiles. Farm tenure does not become very secure, except for the richest quintile. The number of children tends to fall the higher the quintile, as does the number of working age adults in the home. Embeddedness in the market (measured as the proportion of crops sold) tends to rise, and crop diversification tends to fall with per capita expenditure quintiles. The proportion of Spanish speakers tends to rise very steadily (with an especially big jump between 1st and 2nd quintiles) and – no doubt linked – the proportion of communities which practise voluntary unpaid exchange of agricultural labour tends to fall.

Overall, the picture is one where farm assets and education are associated with being in the highest quintile, as is market embeddedness and being Spanish speaking. Both a high off-farm income share and a high degree of crop diversification are associated with being poorer.

Table 2-9: Summary statistics by region

	Mean		
	Sierra	Coast	Rainforest
Total expenditure (soles)	4,243.01	6,677.63	4,586.77
Total expenditure per capita (soles)	1,092.18	1,624.90	1,153.91
Share of family farm income	0.59	0.68	0.56
Share of farm wages	0.11	0.14	0.14
Share of manufacturing self-employment	0.06	0.05	0.12
Share of manufacturing wages	0.08	0.05	0.03
Share of services self-employment	0.11	0.05	0.11
Share of services wages	0.06	0.03	0.04

	Mean		
	Sierra	Coast	Rainforest
Off-farm hours	1,268.48	1,562.32	1,344.60
On-farm hours	3,492.25	3,173.76	2,936.43
Age	39.42	38.70	35.78
Gender (1=male)	0.50	0.52	0.53
Years of education	6.75	7.83	6.65
On-farm experience	16.87	11.54	12.47
Farm tenure (higher means less secure)	1.47	1.84	2.50
Farm capital	854.40	1,043.59	388.71
Farm size	1.87	2.88	10.19
Number children (under 15 years)	2.31	2.14	2.82
Number of adult family members (15-65)	2.89	3.22	2.87
Number of people over 65 in the house	0.26	0.21	0.12
Unearned income (irregular)	285.70	323.89	115.72
Coastal region	0.00	1.00	0.00
Rainforest region	0.00	0.00	1.00
Spanish speaking	0.42	0.90	0.74
Crop diversification index	2.36	1.57	2.13
Proportion of crops sold	0.37	0.84	0.56
Existence of informal labour exchange in community	0.68	0.15	0.70
N	553	160	267

Table 2-9 provides basically the same information as in Tables 2.6 and 2.8, but this time regionally differentiated. One can see that nominal incomes in the coastal region are a good deal higher than in the other regions. Off farm shares are lowest there and agriculture is most commercialised. It is noteworthy, however, that off-farm hours are high in the coastal region compared to agricultural hours. Farm capital (partly measured by the value of traction animals) is low in the rain-forest region. The amount of non-Spanish speaking families is far higher in the sierra than elsewhere, and the practice of unpaid labour exchange is also higher there; education is lower, diversification higher and market embeddedness lower.

Table 2-10 below shows the sectors worked in by the 672 people who declare that their main occupation is non-agricultural.

Table 2-10: Hourly wages and self-employment returns in non-agricultural sectors

	Hourly wages	Self-employed returns	Total
	Mean (N)	Mean (N)	N
Professionals	2.17 (54)	0.12 (13)	67
Lower administration	0.90 (12)	0.09 (2)	14
Sales, and Selling Crafts etc.	2.66 (22)	0.03 (183)	205
Other Services	0.77 (52)	0.13 (21)	73
Forestry and Fishing	2.15 (10)	2.44 (36)	46
Textile Workers	1.33 (10)	0.87 (30)	40)
Preparing Food	1.12 (7)	5.17 (10)	17
Tailors etc.	0.83 (2)	1.21 (23)	25
Plumbers, Electricians etc.	1.07 (18)	1.66 (18)	36
Construction	1.06 (90)	1.24 (30)	120
Transport	1.62 (16)	8.71 (13)	29
Total	293	379	672

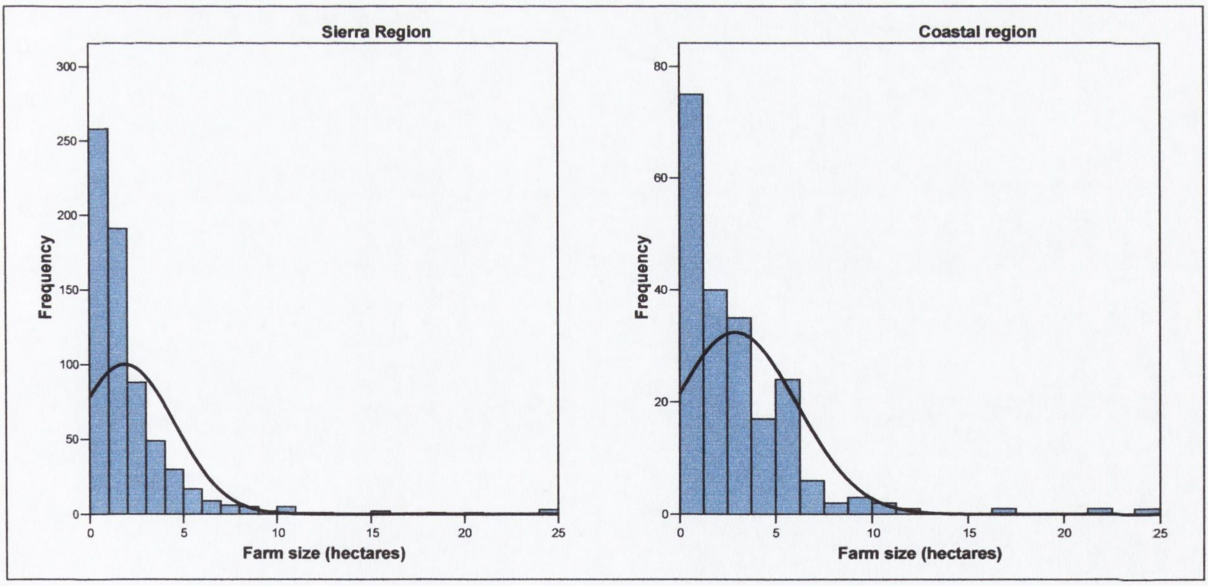
The hourly returns data from self-employment seems particularly unreliable (as noted above also for Table 2-7) and this data is not used anywhere in what follows; the wage data is used in Chapter 5. The hours data for self-employment is used in the hours estimations in the appendix to Chapter 4.

2.2 Visually exploring some relationships

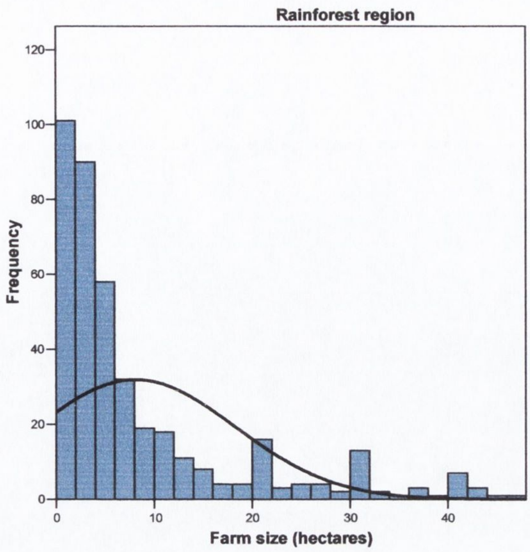
Some particular variables that will play an important part in later chapters include a) off-farm hours, b) farm size, c) total household expenditure (as a proxy for permanent income), d) education (measured as years of schooling), e) the proportion of crops sold (an indicator of market embeddedness) and f) a risk management or diversification index (the inverted Simpson Index) for crops[12]. These variables are now examined visually. There is some repetition of what has just been presented, but in a new medium, which can clarify relationships.

In Figure 2-2 the farm size distribution is shown for each of the three regions xxx(with the normal distribution superimposed)xxx. Farms are smaller in the sierra with a mean of 4.4 hectares and a median of 1. In the coastal region the mean farm size is 2.91, with a median of 2. In the rainforest region, where farms are larger, the mean size is 11.4 hectares, with a median of 4.75.

Figure 2-2: Farm Size in the 3 Regions (normal distribution super-imposed)



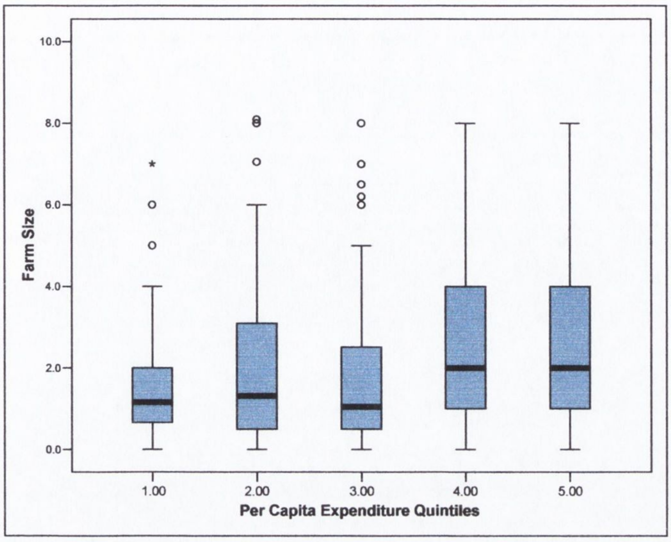
[12] Diversification Index = Inverted Simpson Index = $\frac{q_i^2}{\sum_{s=1}^{n_i} q_{is}^2} \le n_i$, where n_i is the number of different crops harvested by the household i in 1994, q_{is} is the (self-reported) value of crop s and $q_i = \sum_{s=1}^{n_i} q_{is}$.



In xxxFigures 2-3 to 2-6 box plots are used to show a number of bivariate relationships. In each case a numeric variable is plotted on the vertical axis and a nominally or ordinally scaled variable with five values is plotted on the horizontal axis. These plots contain quite a lot of information. They display the median as well as the full four quartiles of the numeric variable at each of the five values of the ordinally or nominally scaled variable, and they also display a number of outliers. In some of the plots extreme outliers have been cut to preserve the main thrust of the visual information (e.g. the few farms above ten hectares have been eliminated). In all the plots the top and bottom numbers on the vertical scale show the maximum and minimum values that are included for the numeric variablexxx.

In Figure 2-3 we see the relationship between farm size and per capita expenditure (where adults are weighted at 1, and children at .5). These two variables are closely related. Per capita expenditure is often taken as a proxy for permanent income, so it seems – unsurprisingly – from this figure that assets, in the form of land, and permanent income, in the form of per capita expenditure, are positively related.

Figure 2-3: Farm size and per capita expenditure



The next three figures present the links between education and farm size, education and off-farm hours and education and per capita expenditure. In the first of these (Figure 2-4) it does appear that there is one class of very small farmers with extremely little education, but otherwise there is little relation between education and farm size (the most educated have smaller farms than those in the 2nd education quintile). In the second (Figure 2-5), there appears to be a reasonably strong, but non-linear, relation between education and off-farm hours. In the third, (Figure 2-6) there seems to be a strong, but fairly linear, relationship between education and per capita expenditure.

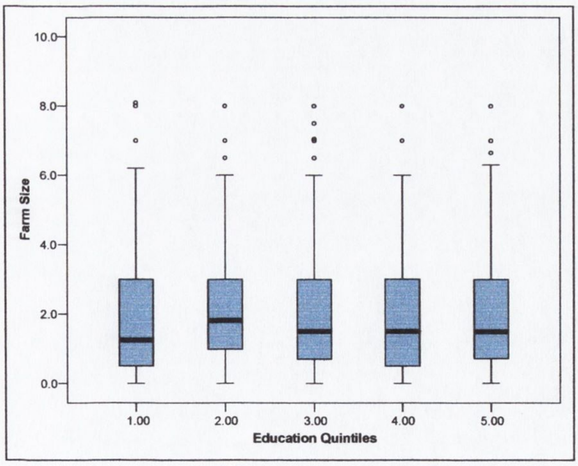


Figure 2-4: Education and Farm Size

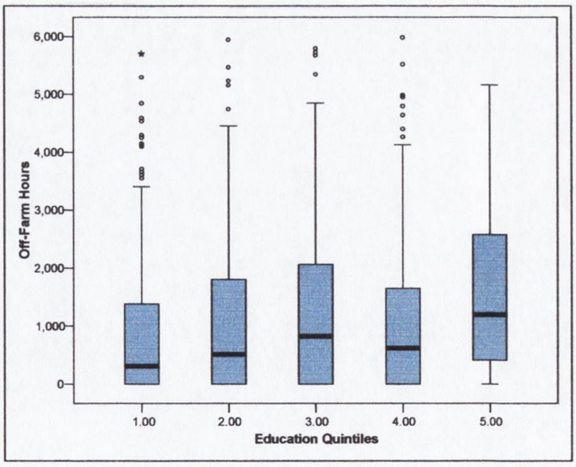


Figure 2-5: Education and Off-Farm Hours

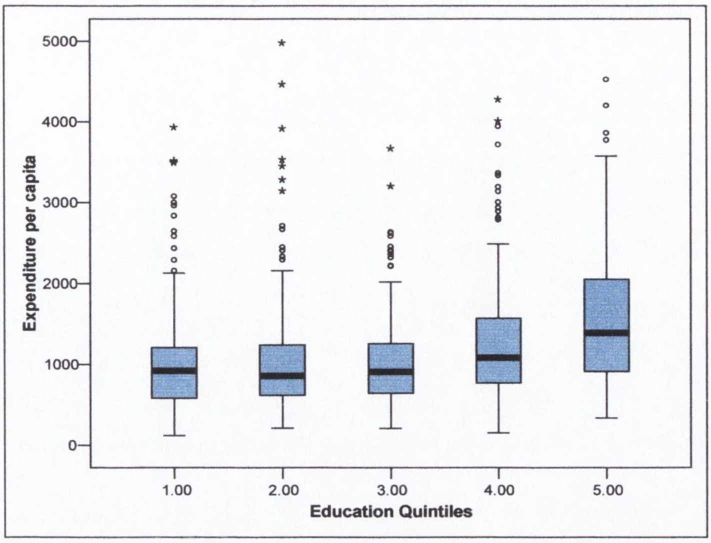


Figure 2-6: Education and per-capita expenditure

In Figure 2-7 we see the relationship between, first, the number of income sources and per-capita expenditure and total expenditure. The relationships are negative and positive respectively.

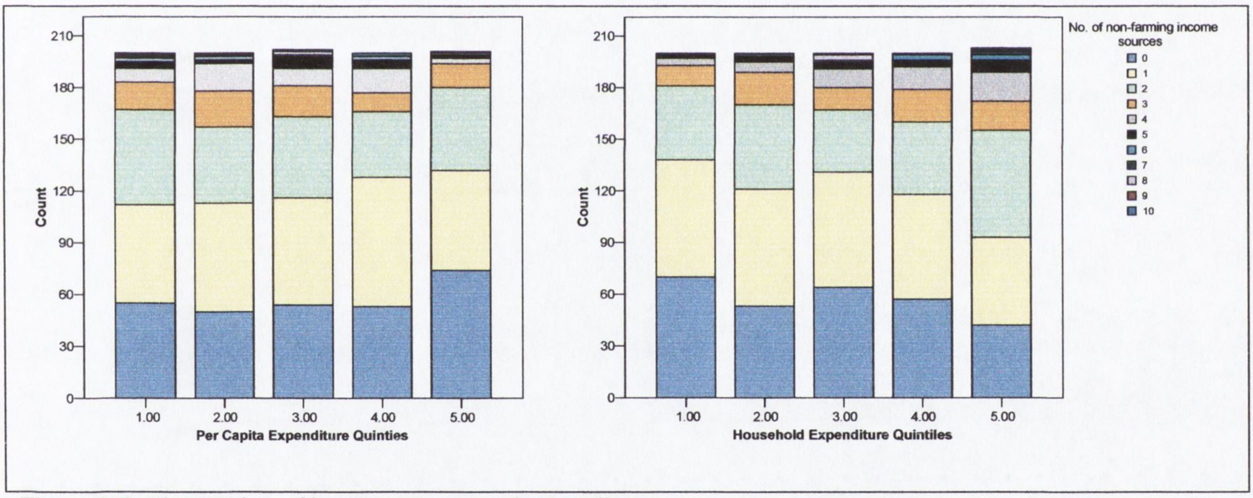
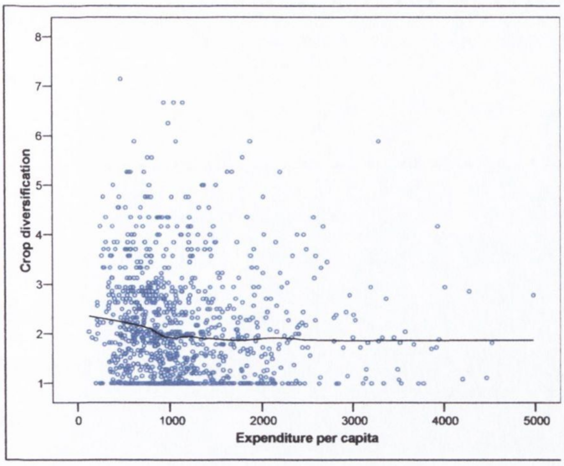


Figure 2-7: The number of income sources and expenditure

In Figures 2-8 and 2-9 (where LOESS lines are drawn in), the relation between the crop diversification index and the proportion of crops sold and per capita expenditure are fairly clear, filling out the information from Table 2-9. Higher diversification is associated with low permanent income. A higher proportion of crops sold is associated with high permanent income.



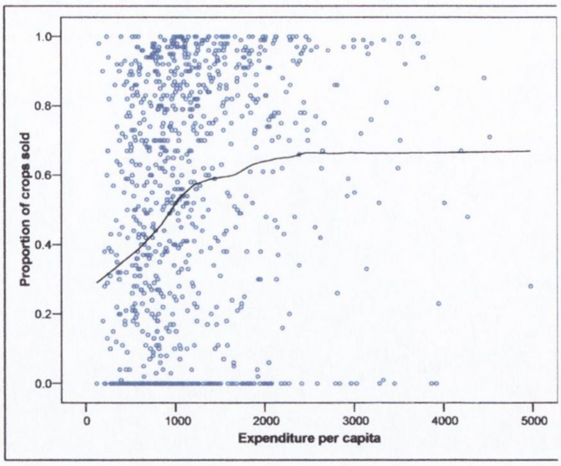


Figure 2-8: Crop diversification and per capita expenditure (LOESS line drawn in)

Figure 2-9: Proportion of crops sold and per capita expenditure (LOESS line drawn in)

Two exploratory regressions in Table 2-11 highlight the relationship between education, farm size and permanent income. The regressions are run for the bottom two permanent income quintiles, and the top two. The first of these groups is very poor. Robles (1997) shows that for 1996 over 60% of households in the rainforest cannot afford a basket of basic goods. The figures are 53% for the sierra and around 40% for the coastal region.

Table 2-11: Exploratory regression on total expenditure (1)

Dependent Variable	**Total expenditure**	
	...for bottom 2 quintiles	**...for top 2 quintiles**
Years of education	50.48** [0.04]	444.06*** [0.00]
Farm size	77.05*** [0.00]	178.28** [0.05]
Farm size * education	-7.90** [0.04]	-12.05 [0.17]
Spanish speaking	506.46*** [0.00]	-113.44 [0.81]
Age	-16.47*** [0.00]	-17.21 [0.32]
Costal region	60.55 [0.80]	1,492.49*** [0.00]
Rainforest region	-158.72 [0.32]	-909.79 [0.11]
Intercept	2,735.51*** [0.00]	3,658.58*** [0.00]
N	392	392
R^2	0.112	0.172

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets.

The land-education interaction term is negative for the bottom two expenditure quintiles. Education, it seems, is more valuable to those on small farms than to those on large farms, and land more valuable to those with little education than to those with a lot of education. For people on mean farm size and education levels for the bottom quintiles 1 extra hectare is more valuable than 1 extra year of education, but when comparing the effects of proportionately equal increases in education and farm size, education has the stronger effect.

Interestingly, in this regard, there is a small (insignificant) negative bivariate correlation between education and farm size in the second poorest quintile (-.05) and a small positive correlation for the top quintile (.1) – and no correlation for the other quintiles. Given that assets and education seem to be routes out of poverty for the extremely poor, for a few households at least they seem to be alternative strategies rather than complementary. This is not quite the case for people at the top of the permanent income scale, for whom extra education becomes very valuable irrespective of farm size, and progressively more valuable than land.

It is notable also that being Spanish speaking is advantageous at the bottom of the income scale, but not at the top (where there are still about 25% of households who do not use Spanish as a first language).

In Table 2-12 below, the same exploratory regressions as in 2-11 are repeated, but crop diversification and off-farm income diversification variables (the number of household income sources outside agriculture) are added. It is very notable that both kinds of diversification appear to be important for poorer people, and not so for the wealthy. The diversification interaction term is negative for the poorer group, but this time insignificant. The education variable is insignificant in Table 2-12 for the poorer group, suggesting that most of its effect in Table 2-11 was through off-farm work. Similar but more fully specified regressions are run at the end of Chapter 6 and the issues further discussed there. But the suggestions that there appear to be two routes out of poverty (land and education) and that these routes are possibly associated with alternative risk management strategies for the poor (on and off-farm diversification) are useful starting points.

Table 2-12: Exploratory regression on total expenditure (2)

Dependent variable	Total expenditure ....for bottom 2 quintiles	....for top 2 quintiles
Years of education	39.77 [0.10]	432.04*** [0.00]
Farm size	74.91*** [0.00]	176.97** [0.05]
Farm size * education	-7.15* [0.06]	-11.33 [0.19]
Spanish speaking	556.56*** [0.00]	-46.65 [0.92]
Age	-12.52** [0.03]	-7.08 [0.69]
Costal region	162.64 [0.48]	1,532.34*** [0.00]
Rainforest region	-173.67 [0.26]	-871.98 [0.13]
Inverted Simpson Index	132.28* [0.08]	99.18 [0.714]
Number of off-farm income sources	337.66*** [0.00]	192.98 [0.60]
Index * sources	-54.90 [0.21]	95.21 [0.55]
Intercept	2,026.50*** [0.00]	2,594.25* [0.07]
N	392	392
R^2	0.173	0.186

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets.

2.3 A typology of off-farm work

Finally, with the aim of informally examining the relationship between individual, household and other characteristics and sector worked in, a K-Means Cluster Analysis of some of the main characteristics at household, farm and regional level reveal that there are five main clusters. The variable loadings for each cluster are shown in Table 2.A.1 in the Appendix.

Summary of Cluster Characteristics:

Cluster 1: Operators in the sierra, quite old, with moderately-sized well-capitalised farms, with moderate crop diversification but highly uncommercial – mainly growing staples. Non-Spanish speaking. (36% of wage earners).

Cluster 2: Younger well-educated females and spouses in large families on small farms in the sierra, with much home consumption. Non-Spanish speaking. (10% of wage earners).

Cluster 3: Relatively poorly educated spouses, in the jungle region, on relatively large undercapitalised but well diversified farms, whose main crop is bananas. Spanish speaking. (7% of wage earners).

Cluster 4: Young, relatively well educated males and females on commercial farms in the coastal area. Spanish speakers. (26% of wage earners).

Cluster 5: Poorly educated operators and spouses in the jungle area on relatively large undercapitalised, uncommercial and undiversified farms. Spanish speaking. (21% of wage earners).

These five clusters are then mapped onto the eleven non-agricultural sectors in the two dimensional Correspondence Analysis below:

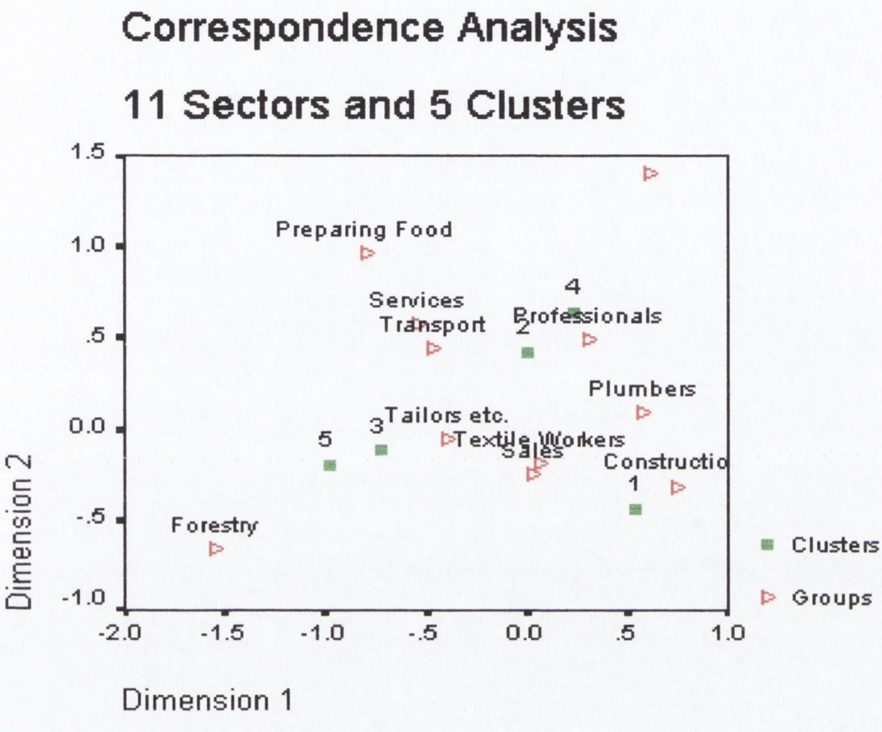


Figure 2-10: Two Dimensional Correspondence between Clusters and Sectors

From this mapping we see that operators in the sierra tend to work outside agriculture in construction and related activities (Cluster 1). They are more likely to work in forestry if they live in the rainforest region (Cluster 5). Professional (mainly teaching) and administrative (mainly secretarial) work is done by young males and females, both in the sierra and on the coast (Clusters 2 and 4 – the administrative group is the un-named icon in the top right of the box). Selling, traditional crafts and industrial work (with wood, metal or textiles) are mainly done by less educated operators or spouses (Clusters 1 and 3) and by non-Spanish speakers (Cluster 1). These mappings correspond to what is known from the sociological and anthropological literature (van der Ploeg, 1990) and provide a simple picture of a fairly complex story.

2.4 Conclusion

This chapter includes both an examination of much of the raw data that is used later in this study, and a brief description of agriculture and farm households in Peru in 1994. Before any formal econometric estimations have been carried out, a number of things appear quite clear, and a number of hypotheses are suggested. To take just a few examples, it is clear that off-farm income is an important part of household incomes, especially of poorer households; it is clear that larger farm size and more education are both associated with increased incomes but that there is little correlation between these two variables themselves; it is clear, also, that while having more income sources is associated with greater total expenditure it is associated with less per capita expenditure (where adults are weighted at twice what a child is weighted); among the hypotheses suggested are those concerning the roles of crop diversification and off-farm work as (substituting or complementary?) risk management practices. In the next chapters (especially chapters 4, 5 and 6), these and related issues are more thoroughly investigated.

Finally, a comment on the variables themselves: detailed variable descriptions have not been included in this chapter, just enough information to (hopefully) ensure comprehensibility. A detailed description of the variables, and discussion of some measurement issues are relegated to the appendix of Chapter 4. The main reason for placing them in Chapter 4 is that most of the variables are used for the first time in that chapter (only a small subset is used in Chapter 3). The main reason for placing them in an appendix rather than in the main text of this chapter (or Chapter 4) is to allow space for discussion and definitions. The number of variables involved is quite large and it was felt that including the variable list and ensuing discussion in the main text might bog it down more than was necessary.

3

Collective versus unitary model

Introduction

In this chapter, a formal test of the Collective Model, involving the estimation of a demand system and tests using the properties of the system, is carried out. The test is carried out using consumption data. The main hypothesis tested is whether the Collective model or the unitary model is appropriate for rural Peru. Supplementary tests include a new test for three adult families (quite common in rural areas).

The sample is later split to repeat tests for Spanish speaking and non-Spanish speaking groups. This test has been motivated by the observation, noted in the last chapter (and made by Deere and Leon, 1982), that couples in indigenous and Spanish speaking households tend to have different interaction patterns. In most indigenous families financial decisions tend to be a jointly made, whereas in Spanish or mestizo families most financial decisions are made mainly by the male. A further test, among Spanish households only, is carried out to test if the same family model holds for those households more or less exposed to the market. This test is motivated by the possibility that exposure to the market changes the nature of intra-household interactions.

3.1 Testing for the collective model

Although the main overall focus of this dissertation is on labour supply choices, the tests in this chapter use consumption data rather than labour supply data. This is simply because distribution factors are not available for the tests to be carried out with labour supply data.

If there is sufficient price variability testing for the Collective Model is possible using consumption data by examining the rank of the adjusted Slutsky matrix. If there is no price variation, then, as seen in Chapter 1, either distribution factors or exclusive or assignable consumption items must be found (e.g. men's or women's clothing). The latter are not available in the data set (except for the money spent by each person on eating out – and not many do eat out) but price data is (from about half of the survey segmentos), so the Slutsky matrix approach is the one used.

Probably the clearest explanation of the logic of this test appears in Vermeulen (2002) though it was first formulated in Browning and Chiappori (1998). In the Unitary model, as Vermeulen describes, familiar Slutsky effects capture the move along an indifference curve as price changes. However, in the Collective Model, there is not only a move along the indifference curve of the household to consider but the movement of the curve itself. This movement occurs because, when we hold household income constant after a price change, the bargaining weight – μ – is not held constant. A price change affects the position of household indifference curve through its affects on the bargaining power of the partners.

If we let $S = \Sigma + uv'$, (3 – 1)

then this may be called the pseudo-Slutsky matrix, with transpose S'. Σ is the symmetric Slutsky matrix of the Unitary model and the accompanying (rank one) matrix formed by multiplication of the two vectors u and v is a measure of the effect of changed bargaining power on behaviour. The vector v measures the effect of price on bargaining weights – $v_j = \frac{\partial \mu}{\partial p_j} + \frac{\partial \mu}{\partial x}\xi_j$ – (3 – 2)

and u measures the effects of changed weights on behaviour – $u_i = \frac{\partial f_i}{\partial \mu}$, (3 – 3)

where $\xi(p,x) = f(p,x,\mu(p,x))$ (3 – 4)

is the observed demand function.

Given that M, where $M = S - S'$, can be shown to be an antisymmetric matrix and that a real antisymmetric matrix must have even rank, Browning and Chiappori (1998) show that M will have a rank of at most 2, whereas the symmetric Slutsky matrix Σ has a rank of zero. Browning and Chiappori go on to show that M has rank 2 if and only if

$m_{ij} = (m_{1i}m_{2j} - m_{1j}m_{2i}) / m_{12}$ for all i,j such that $j > i > 2$. (3 – 5)

To carry out a test of the rank of M one needs price and consumption information on at least 5 goods. If there are exactly five goods in the system, then i =3 and j = 4 and there is only a single restriction. The (arbitrary) fifth good is necessary to ensure adding up.

The restriction to be tested if there are potentially three decision makers (and so three utility weights in the household utility function) is more complicated. Working through the algebra, one finds that in a 7 good system it is the following:

$$m_{56}m_{12} = -m_{16}m_{25} + m_{16}m_{23}[F] + m_{16}m_{24}[G] + m_{26}m_{15} - m_{26}m_{13}[F] - m_{26}m_{14}[G] + m_{36}m_{12}[F] + m_{46}m_{12}[G],$$

$$where \quad F = \frac{m_{14}m_{25} - m_{24}m_{15} - m_{45}m_{12}}{m_{23}m_{14} - m_{24}m_{13} + m_{34}m_{12}} \quad and \quad G = \frac{m_{25}m_{13} - m_{23}m_{15} - m_{35}m_{12}}{m_{24}m_{13} - m_{23}m_{14} - m_{34}m_{12}} \qquad (3-6)$$

Again, the seventh good is needed for adding up to hold (the basis for the result above is reported in the appendix to this chapter).

The null of this test is the general version of the Collective model (where personal utility depends on both own and spouse's – or other family members' – consumption and leisure, and not on their utility), and rejection of the test therefore implies a rejection of this relatively unrestricted version of the Collective model. The most likely reasons for such a rejection, should it occur, include a) the possibility that a non-cooperative game among family members leads to a non-Pareto optimal outcome, b) the possibility that the outcomes are constrained by traditional gender (or parental) roles within the family and/or c) the possibility that bargaining weights are not affected by prices as postulated by the model as introduced in Chapter 1. If there is rejection, therefore, further investigation is needed as to its cause.

3.2 Model specifications

A 7 good AIDS demand system is estimated for rural Peru. Price information is available for (at least) six specific goods that are commonly reported in people's baskets (rice, cleaning products – soap and detergent, sugar, mineral water and soft drinks, meat and cooking oil) and a seventh aggregate good is implied, comprised of the rest of food consumption. Food consumption (plus soap) is assumed separable from other consumption. The price information used in the estimations is information on prices for a number of foodstuffs and other basic goods given by community heads (for their own community only) in response to a separate series of questions given together with Living Standards Measurement Household Survey. This information is only as good as the community leaders' memory and knowledge will allow. Nevertheless, it is unusual to have such geographically varied price data on basic foodstuffs (as opposed to data on expenditure or receipts). Initially, it was intended that only households who do not 'autoconsume' any of these goods would be included in the sample to avoid conflation of income effects, but the sample size declined too drastically. The compromise reached was that households who buy in their entire consumption of at least 5 of the 6 named products would be included in the sample. The sample was also reduced due to missing price data in a number of the segmentos and to there not being two partners in a number of households. The reduction in potential sample size from the 980 households of Chapters 2, 4, 5 and 6 leaves a sample of 499 households.

The AIDS approach is common in the literature and is derived from a utility function specified as a second order approximation of any arbitrary utility function. The AIDS equations satisfy the axioms of choice and, unlike the Linear Demand System, can be aggregated over consumers without imposing parallel Engel curves. Deaton and Muellbauer (1980) start from a cost function which allows exact aggregation over consumers. If u represents utility and p prices, then one class of exactly aggregable cost function can be represented as:

$$\ln c(u,p) = (1-u)\ln\{a(p)\} + u\ln\{b(p)\} \qquad (3-7)$$

where a and b are functions of prices and can take the following specific flexible functional forms:

$$\ln a(p) = \alpha_0 + \sum_i \alpha_i \ln p_i + .5 * \sum_i \sum_j \gamma_{ij}^* \ln p_i \ln p_j \qquad (3-8)$$

and $\ln b(p) = \ln a(p) + \beta_0 \Pi_i \, p_i^{\beta_{ik}}$. $\qquad (3-9)$

Substituting both of these into the cost function gives:

$$\ln c(u,p) = \alpha_0 + \sum_i \alpha_i \ln p_i + .5 * \sum_i \sum_j \gamma_{ij}^* \ln p_i \ln p_j + u\beta_0 \prod p_i^{\beta_i} \qquad (3-10)$$

The logarithmic derivative of the cost function with respect to price gives the budget share equation: $s_i = \sum_{k=0}^{K} \alpha_{ik} X_k + \sum_{j=1}^{n} \gamma_{ij} \ln p_j + \beta_i \ln(E/P) + e_i$, $\qquad (3-11)$

where $\gamma_{ij} = .5*(\gamma^*_{ij} + \gamma^*_{ji})$ and the i subscript represents each of the six goods whose shares (s) are estimated, j represents the six goods in each estimated equation, k indexes factors affecting household preferences (X), and E/P is real expenditure (total food expenditure divided by a price index[13]). Homogeneity and adding-up are imposed, while symmetry can be tested for (γ_{ij} = γ_{ji}, yielding 15 restrictions). The adjusted Slutsky matrix predicted by the Collective Model has restrictions that are highly non-linear. The tests based on this matrix will be discussed as they arise below but homogeneity and adding up are imposed in all the regressions of this chapter.

There are several practical difficulties that emerge in this kind of estimation. One arises when, as in this case, there are a significant amount of zero shares on the left hand side. There are at the time of writing at least five ways of dealing with censoring in a systems context. Two step methods have been proposed by Heien and Wessels (1990) and by Shonkwiler and Yen (1999). The former have been shown to be inconsistent (Shonkwiler and Yen, 1999). The latter can also be estimated as a single step. Single step methods have also been proposed – simulated maximum likelihood (Kao *et al.*, 2001), quasi-maximum likelihood (Yen *et al.*, 2003) and an information theoretic approach (Golan *et al.* 2001). In this paper, one two-step – the consistent two-step approach of Shonkwiler and Yen – and one single step method – the information theoretic approach by Golan *et al.* – are used.

The system proposed by Shonkwiler and Yen as adapted here means estimating the following: $y_{it} = \Phi(z'_{it}\alpha_i) f(c'_{it}\beta_i) + \delta_i \phi(z'_{it}\alpha_i) + \varepsilon_{it}$, (3 – 12)

where z are the factors affecting participation and c are the factors affecting consumption – prices, household characteristics etc – and y are the shares, and the two functions are, respectively, the cumulative standard normal and the standard normal density. The subscript t indexes observations. The fitted cumulative densities and the fitted densities themselves come from probit estimations carried out in step 1, and the

[13] If we substitute E for c in the cost function and solve with regard to u, we get the indirect utility function, which can then be used to substitute for u in the budget share equation. The price index is sometimes assumed to be a Stone price index for AIDS models, which renders the model linear and estimation easier. In this case an iterative process that slightly simplifies Browning and Meghir, 1991, is used. P= $\sum_{i=1}^{6} \rho_{i1} \ln(price_i) + .5\sum_{i=1}^{36} \gamma_{ij} \ln price_i \ln price_j$ is calculated after each Shonkwiler/Yen estimation and this substitutes for P in the next estimation and the process is repeated until changes are minimal (3 or 4 iterations in practice). The same index is used later in the Entropy model. The initial index for the first iteration is the Stone index. All prices are relative to the price of the seventh good – other food consumption, which is the average 'segmento' price of the other priced food items in the survey.

equation above using a SUR estimation for all six goods is carried out in the second step. The approach is used here primarily as an alternative model to compare with the information theoretic procedure, which needs few distributional assumptions and tends to have good small sample properties (Golan *et al.* 1996). The highly non-linear test for the three person version of the Collective Model is easier to do using the information theoretic approach, while the two-step approach is used both to help in constructing the information theoretic estimation (as shall be seen below) and in gauging its plausibility when comparing common results for tests of Unitary and two-person Collective restrictions.

In particular, the single step information theoretic approach of Golan *et al.* (1996) uses a maximum entropy estimator (Golan *et al.* 1996; Mittlehammer *et al.* 2000) with weak distributional assumptions. The following equations are estimated here:

$$s_{it} = \sum_{k=o}^{K}\sum_{d=1}^{D} z_{ikd}^{\rho} q_{ikd}^{\rho} x_{tk} + \sum_{j=1}^{n}\sum_{d=1}^{D} z_{ijd}^{\gamma} q_{ijd}^{\gamma} \ln(p_{ij}) + \sum_{d=1}^{D} z_{d}^{\beta} q_{id}^{\beta} \ln(\frac{E_t}{P_t}) + \sum_{h=1}^{H} v_h w_{ith}, \quad (3-13)$$

when shares are greater than zero, and

$$s_{it} > \sum_{k=o}^{K}\sum_{d=1}^{D} z_{ikd}^{\rho} q_{ikd}^{\rho} x_{tk} + \sum_{j=1}^{n}\sum_{d=1}^{D} z_{ijd}^{\gamma} q_{ijd}^{\gamma} \ln(p_{ij}) + \sum_{d=1}^{D} z_{d}^{\beta} q_{id}^{\beta} \ln(\frac{E_t}{P_t}) + \sum_{h=1}^{H} v_h w_{ith}, \quad (3-14)$$

when shares are equal to zero (no household consumption of the product takes place). In the equations, D and H are the number of support points (z and v) for, respectively, the main probabilities (q) and the probabilities used in calculating the error term (w). In each case there are two, and all pairs of probabilities are constrained to add up to 1.

Testing for symmetry means testing that $\sum_{d=1}^{D} z_{ijd}^{\gamma} q_{ijd}^{\gamma} = \sum_{d=1}^{D} z_{jid}^{\gamma} q_{jid}^{\gamma}$, (3 – 15)

and again this means there are 15 restrictions.

Two important practical difficulties occur when using entropy estimators – firstly, how to weight the known and the unknown parts of the model (the data terms and the error terms) and, secondly, how to decide on the supports (z and v). The first question is answered by weighting the errors so that the chi-squared values for the symmetry test using the entropy estimator are of the same order of magnitude as those when using the two-step estimator. The result is that error observations have weights of 1000 compared to the data point observations (the iterative process involved gave error weights of .00001, .0001.,.001,.01,.1,1,10,100,1000,10000). The second question is answered by using supports that are +120, -120 for prices and X variables, and +100,-100 for the error terms. These were the lowest symmetric values compatible with general feasibility for *all* the different versions of the estimations run, although much lower values were possible for the vast majority of the estimations (the instrumental variable model below, in particular, necessitated wide supports). The main effects of a re-weighting downwards of the errors (from the reported 1000 to 100, 10 and then 1 and lower) is that for the very low weights no hypothesis whatsoever is ever rejected – though it is important to note that the various calculated chi-square statistics maintain very close to the same ratios for the unitary and collective tests for all weights as those quoted below. The implication of all of this is that any rejection reported from the entropy estimations of this section is partly due to the relatively high premium accorded to signal accuracy in the estimations, and the relatively low premium to robustness. The high weighting of the errors, that is, affects statistical significance decisions, but does not affect the relative effects of the various constraints on the objective. There is no significant effect on the main test results from using a range of different support points for those estimations where narrower supports are possible.

In a seven good system there are, in fact, six restrictions for the two-adult Collective model, which can be summarized as follows: $m_{ij} = (m_{1i}m_{2j} - m_{1j}m_{2i}) / m_{12}$, where the six i,j combinations are 3,4; 3,5; 3,6; 4,5; 4,6 and 5,6. m_{ij} is the difference between the i coefficient and the j coefficient in the 6*6 matrix of price coefficients from the system (e.g. m_{12} is the price coefficient for good 2 in the equation for good 1 minus the price coefficient for good 1 in the equation for good 2: $\sum_{d=1}^{D} z_{12d}^{\gamma} q_{12d}^{\gamma} - \sum_{d=1}^{D} z_{21d}^{\gamma} q_{21d}^{\gamma}$).

The actual test statistic is chi-squared, and is twice the difference of the unrestricted and unrestricted objective functions (closely analogous to a likelihood ratio test).

These are the same kinds of restrictions that need to be imposed for the Shonkwiler and Yen estimations, except that the m's used in the joint test are now differences in adjusted coefficients ($\Phi(z_{it}^{'}\alpha_i)\gamma_{ij} - \Phi(z'_{jt}\,\alpha_i)\gamma_{ji}$, from Alasia and Soregaroli, 2002) rather than differences in the coefficients themselves. In the Shonkwiler and Yen model, Wald tests are carried out on the unrestricted equations. In the present form of the test the SUR standard errors are used.

In most AIDS estimations the expenditure variable is modelled as an endogenous variable, instrumented usually by income. In this case, income and income squared are used as instruments in a second set of entropy equations, and solved simultaneously for the three versions of the model – the unrestricted version (actually imposing adding up and homogeneity), the version further restricted by symmetry (15 unitary model restrictions), and the third version, restricted by the 6 Collective Restrictions.

One final and important question concerns the possibility of testing for there being more than two decision makers within a Collective Framework. It is proposed that this test be carried out using the Entropy model.[14] To motivate this initially, the sample is broken up to check the plausibility of there being more than two decision makers. Three samples are chosen – all 499 households; the 212 households which have a couple, but no adult children; and the 287 households with a couple and one or more adult children. If the Collective Model does appear to validly represent behaviour, and if all adults contribute to the decision making process (Fafchamps and Quisumbing, 2000) then the restrictions of the Collective Model should not be rejected for a couple without adult children, but should be rejected when there is one adult child, let alone more.

[14] The single restriction for the three person model is stated in Section 3.3.

3.3 Results

The chi-squared statistics from the restrictions in the two-step model are as follows:

Whole Sample: Unitary Restrictions (UR) = 135. Collective Restrictions (CR) =34.7.

Couples without Adult Children: UR 48.7. CR 15.02.

Since 25 is the 95% critical value for the 15 degrees of freedom of the unitary test, and 12.6 is the 95% critical value for the 6 degrees of freedom of the collective test (16.8 is the 99% value) it can be seen that all the restrictions are rejected at 5 % though the Collective Model restriction in households with a couple, but without adult children, is not rejected at 1% (p-value=.02).

Turning now to the information-theoretic model, when expenditure is uninstrumented, we find the following very similar chi-squared results:

Whole Sample: UR 117.25 CR 22.12

Couples without Adult Children: UR 60.51 CR. 12.4

Once again, symmetry is rejected. The collective restrictions are strongly rejected for the couple with adult children, but not for couples without children. The similarity of all the information-theoretic results to the results using the conventional estimator, when only the first test of the four was used as a yardstick in the weight selection process, gives added confidence to the information theoretic results for the more complex three person test below.

When expenditure is instrumented by income and income squared, the chi-squared results for the entropy model are as follows:

Whole Sample: UR 370.02. CR 74.8

Couples without Adult Children: UR 17.69. CR..074

Both sets of restrictions are rejected for the whole sample. Neither set are rejected for the reduced sample, but the p-value is higher for the Collective Restrictions.

Calculating results from very distinct methodologies, as is done here, is always likely to leave hostages to fortune, and judgement must be used. With regard to the question of whether expenditure should, as is usual, be instrumented, separate endogeneity tests and over-identifying tests on the six regressions in a conventional (non-censored) framework suggests probably not. How valid are these tests? The number of zeroes, out of 499, for the 6 products estimated are 65 for rice, 24 for cleaning products, 68 for sugar, 343 for mineral water, 276 for meat and 53 for oil, so the degree of censoring varies quite a bit. Certainly for the first three, and the last, the tests are probably reasonably reliable. Of those four the null of exogeneity is rejected only in the cleaning products estimations, and then only at 10%, not at 5%. The p-values for the six Wu-Hausman endogeneity tests, for the six products above, respectively, are: .52, .07, .99, .11, .25 and .36. The Sargan over-identifying restrictions are only rejected in the meat estimation. Their p-values are: .31, .22, .32, .33, .02, .17. These results suggest that income and income squared probably are good instruments, as one might expect. The Wu-Hausman results, on the other hand, suggest they are unnecessary.

Since the endogeneity results suggest favouring the chi-squared statistics from the non-instrumented regressions then symmetry, and hence the Unitary Model, is clearly rejected and the Collective Model is also rejected in couples with adult children, but with mixed results for couples with no adult children.

With regard to the more complex test for 3 decision makers in larger families, we find – first – that in families with just one adult child (of which there are only 112), the Unitary model is strongly rejected. The chi-squared statistic is 36.2 for the fifteen Unitary symmetry restrictions. For the single restriction of the 3-person Collective Model, the chi-squared statistic is .45 (the critical value is 3.84). This is strongly suggestive of an active role in family decision making for the 'adult child'. If we turn to families with more than one adult child, the chi-squared statistic is 7.04 for the same three-person restriction for the 176 families with more than one adult child. This means rejection, even at 1% (the 1% critical value with just one degree of freedom is 6.64).

The results of these information theoretic models, allowing for the relatively small samples, are generally supportive of the Collective Model, where the number of decision makers corresponds with the number of adults in each family. The Unitary Model is always rejected. The two person Collective Model is not rejected for two adult person families (or only barely rejected, at 5%, for the conventional estimator), but is when families have more than two adults. The three person version of the model is not rejected for families with three adults but is, in turn, for larger families.

As noted in the introduction to this chapter, some of the anthropological literature posits different types of intra-family interaction for indigenous and non-indigenous households. It was decided, therefore, to test for the Unitary model among Spanish speaking and non-Spanish households. Given the sample sizes (343 Spanish speaking households and only 156 non-Spanish speaking households) it was only possible to do this for the whole sample for non-Spanish speakers. Using the two-step estimator, symmetry is clearly rejected for the both groups (chi-square statistics were 210.9 for Spanish speakers and 91.42 for non-Spanish speakers). When the sample is reduced for Spanish speakers (into families with 2 adults only), symmetry continues to be rejected (chi-square=102.2), and the Collective model is also rejected (chi-square=57.39). So, even though the Collective model is on the borderline of acceptance for the population as a whole for two-parent only families (there is certainly more evidence against the Unitary than the Collective model) it does seem as if Pareto efficiency for those families is far less likely for Spanish speaking two-parent families than for non-Spanish speaking ones. This result is only suggestive of ethnic differences in family interaction patterns, but it is strongly so. Unfortunately, there is not a large enough sample of two-adult families to be more certain.

One final breakdown of the sample was carried out. The Spanish speaking sample was broken into those above and below median levels of market embeddedness in order to see if differing degrees of exposure to the market appeared to affect the underlying utility model. The two symmetry tests gave almost the same chi-square results (chi-square = 100.8 for those exposed to the market and 119.9 for those not), meaning the Unitary model is rejected in both cases and suggesting that market exposure has probably not fundamentally changed the model of intra-familial decision making among Spanish speakers.

With regard to other results from the demand system estimated: the Hicks elasticity estimates below, for the whole sample, show a reasonably close correspondence for the own price elasticities from the two-step and non-instrumented entropy estimations. Cross-elasticities are somewhat more varied – some of them are quite high in the two step procedure. The instrumented elasticity is slightly above zero for rice, and this casts further doubt on the instrumental variable specification.

Finally, the real expenditure coefficient is positive significant for meat, and negative significant for sugar in the two-step specification, and insignificant for the other four goods. In the entropy specification it is positive for meat and negative for rice, sugar and oil and insignificant for the other variables. A positive expenditure elasticity in an AIDS model indicates an income elasticity greater than 1.

Table 3-1: 3 sets of Hicks elasticities

	Price (ln)					
	Rice	Sugar	Meat	Water	Soap	Oil
Entropy Model (expenditure exogenous)						
Rice	-2.141	0.848	0.140	-0.029	-0.289	0.343
Sugar	-0.611	-0.298	0.237	-0.818	0.533	0.316
Meat	-0.454	0.768	-0.740	-0.520	1.233	-0.255
Water	0.203	-1.656	0.218	-1.645	0.713	0.509
Soap	0.193	-0.122	-0.048	-0.803	-0.695	0.311
Oil	-0.757	0.844	0.238	0.203	0.148	-0.751
Entropy model (expenditure endogenous)						
Rice	0.064	-3.030	-0.278	1.398	-0.131	0.425
Sugar	0.982	-3.090	-0.060	0.220	0.650	0.367
Meat	-3.198	5.644	-0.206	-2.266	1.031	-0.425
Water	-3.440	4.847	0.943	-3.939	0.450	0.271
Soap	2.333	-3.885	-0.456	0.566	-0.458	0.329
Oil	1.397	-2.975	-0.179	1.575	0.303	-0.632
Two-step model						
Rice	-0.911	-0.422	-0.636	1.124	-0.132	0.460
Sugar	-0.094	-0.765	0.389	-0.748	0.060	0.247
Meat	-0.610	0.840	-0.887	-0.414	-0.006	0.413
Water	0.221	0.841	-2.230	-2.067	0.276	1.434
Soap	1.789	-0.322	-2.839	2.011	-1.015	-0.059
Oil	0.766	0.749	-2.882	2.185	-0.548	-1.039

The parameters, and standard errors, on which these elasticities are based are reported in the Appendix.

3.4 Conclusion

This chapter shows that the Collective model is probably the appropriate utility model for farm households in Peru, at least for those households for whom Spanish is not the main language. Both estimators, a conventional two-stage estimator and the Maximum Entropy information-theoretic estimator, give almost the same results for the Unitary tests (rejection) and for the two-person Collective restrictions (rejected in families with adult children, but not – albeit at 1% with the conventional estimator – for two person families). For the first time in the literature the test for the Collective model using a 3-person adjusted Slutsky condition is carried out. This is rejected, using the information theoretic estimator, for families with 2 main persons and 2 or more other adults, but not for those families with 2 main persons and only 1 other adult. This result provides further evidence in favour of the Collective model.

However it does appear as if the Collective model is probably not appropriate for Spanish speaking households, and the relatively favourable (to the Collective model) test results for the country as a whole come from a weighted average of two utility models. This could have important policy implications. If intra-familial shares are decided using a different mechanism in different cultures then the implications for the welfare of individuals within the household of an increase in, say, crop productivity (in regions, say, where crops are generally under the control of the male) are likely to differ for different ethnic or language groups.

Finally, when the Spanish speaking part of the sample is divided into those who are more and less exposed to agricultural markets one finds no difference in the tests results (the Unitary model is rejected for Spanish speakers), suggesting that exposure to markets has probably not affected the underlying household utility model for Spanish speakers.

If the Unitary model does not hold for most households, as seems to be the case, then it is even more important to look inside the black box of the household. Apart from this particular chapter (which, necessarily, uses consumption data), the main focus of this thesis is on off-farm work. In accordance with the findings of the chapter, therefore, the next chapter looks at progressively more individualised and disaggregated models of off-farm work. The chapter after that looks at constraints on off-farm work and on hiring in, first at the household, but then the individual level. Only in Chapter 6, which examines effects of household off-farm strategies, does the focus of attention return to the household as a unit.

4

Reduced form participation models

Introduction

The results of the last chapter suggest that, at least with regard to consumption, the decision making power of individuals is not subsumed into the farm household. Rather, family members make their own decisions (with the proviso, if the Collective model is appropriate, that others are not hurt by the decisions they make). In this chapter, off-farm participation estimations are carried out at three different levels of analysis– the household, the couple and the individual adults of the family. It is expected that by tracking carefully the accretion of detailed results one can reach conclusions about intra-familial decision making, even in reduced form specifications. To this end, a range of discrete choice models of participation is employed. Two adaptations new to the literature are estimated (a multivariate probit and an extended version of the multinomial logit) to model participation of more than two adults, and one relatively new adaptation (a multinomial logit with cross-category sample selection mechanisms) is used to model hours worked in a variety of off-farm categories.

The reported results are quite extensive and fairly detailed, so two sections of the chapter are relegated to the Appendix: first, a section introducing the variables used, with some additional discussion of measurement issues and, second, a section reporting the results of the hours estimations (which add very little to the overall sense of the participation results). What is included in the main body of the text is an introductory section on possible reduced forms, detailed reports of participation model results, and a discussion of the overall results.

The models themselves become progressively more disaggregated throughout the chapter. The number of discrete options is increased (from just 2 - on/off-farm work - to 6 categories of work, including unpaid or paid on-farm work, paid off-farm and non-farm work, and two self-employment categories). As part of the disaggregation, the household itself is disaggregated in the models (from whole household models, to 'couples only' models to models which include all adult family members). The reason for this progressive disaggregation is simply to mark out at each step what the gain from disaggregation is (if there is one), both in statistical terms (model fit) and, more importantly, in terms of the quality and interest of the results. Much of the published literature in the farm household model tradition is carried out at the whole household or couple level, with few discrete options. The less formally theory-based tradition of modelling associated with Reardon et al. (2001) and Ellis (2001), focusing on rural livelihoods, tends to have more discrete labour options or income sources (Reardon et. al., 2001, suggest that the breakdown might follow International Labour Organization categories, the route followed in this chapter), but also usually models at the household or couple level. In this chapter, the main aim is to show that some of the most interesting results can come from the greatest degree of disaggregation of both choices and households.

4.1 Reduced form models of participation

The main reason for investigating a variety of specifications of off-farm labour models with reduced forms[15] is that these are relatively easy to compare when a variety of specifications are used. The amount of choices that have to be made in estimating more structural models would make the kind of step by step comparative exercise carried out in this chapter extremely difficult. As it is, there is a number of choices to be made, and even the reduced form models that are estimated contain variables that may well be regarded as endogenous either within the single period static model of off-farm labour supply, outlined in Chapter 1, or certainly within a life-cycle multi-period labour supply model (Heckman and McCurdy, 1980; McCurdy, 1981 and 1983; Browning *et al.* 1985, and Blundell and Walker, 1986).

Several possible versions of intra-household decision-making present themselves as potential frameworks for reduced farm modelling of off-farm or non-agricultural participation and hours worked off the farm: i) assume that everyone maximizes their own utility, irrespective of what is done by others in the household; ii) the household could simply be treated as a single unit (with the assumption of a joint utility function) and a participation equation and hours equation estimated at the level of the

[15] 'Reduced form' in this case means a very 'pared down' reduced form: for example, two variables that are commonly estimated for inclusion into reduced form models – farm profits and off-farm returns – are not included here, only the (more) exogenous variables which determine them.

household; iii) the individual members of a multi-person household (usually, in practice, 2 people, but potentially more) can be modelled as having a joint Unitary utility function – in which case either a binomial probit or multinomial logit estimation procedure would be appropriate; iv) a Collective utility model may be deemed appropriate, again in a two (or more) person household (although, in a reduced form model, this may not be easily distinguishable from iii).

A fifth option that may be used is v) the recursive 'traditional family' model, where separate regressions are run for male and female family members (male farmers and their spouses); in the first of these regressions, males are treated simply as individuals, whose decisions depend on household unearned income, on their own preferences and their own returns; in the second, females are treated as individuals whose decisions depend on decisions already made by the male, as well as on their own preferences and returns and household unearned income.

A difficulty with estimating iii) and iv) above is that intra-family interactions where more than just operator and spouse are involved are not straightforward to model, though they may be important. So far, only three attempts have been made in the literature to model full intra-family interactions in on-farm and off-farm work decisions (although research into child labour in developing countries usually includes the role of the whole family, intra-family interactions are not typically modelled – Bhalotra (2000) and Bhalotra and Heady (2000)). These have been the works by Newman and Gertler (1994), Fafchamps and Quisumbing (2000) and Malchow-Møller and Svarer (2002) already discussed in Chapter 1, in section 1.1.5. In each of these cases the Unitary household model is assumed.

All the reduced form models of individual household members' participation that are actually estimated below are derivable from one or other version of the farm household model, and include one-person, two-person and multi-person variations of ii), iii) and iv) above. Both i) and v) have also been used in exploratory work, but only a few pertinent results are mentioned.

Specifically, the off-farm participation decision is modelled below at three different levels corresponding to three different units of analysis: model a) at the *household level*, model b) at the *level of operator and spouse* and model c) for *all the (working) adults (over 14s) in the household.* For model a), the set of exogenous variables from

Equation (1-5) are included on the right-hand side. For model b), the relevant equation to be estimated is Equation (1-11).[16] For model c), the approach of Malchow-Møller and Svarer to operationalising Equation (1-13), summarized in Chapter 1, is used for households with more than two people. Each of these simple participation models can, therefore, be motivated directly by theory. In the first 2 cases, the theory comes from the farm household tradition. Statistically, in this tradition, the participation models are seen as random utility models, and a positive value for the dependent variable indicates that the actual wage is higher than the reservation wage, and hence the utility from participation is higher than that from non-participation. The third set of models is a little different and relies for theoretical justification on the somewhat novel approach of Malchow-Møller and Svarer, summarized in Chapter 1. In this case the latent variable is not utility, but desired labour supply.

Furthermore, in the case of model c), the estimations are carried out over three progressively more disaggregated steps:

1) In the first step, the set of estimations have dichotomous dependent variables, where 'one' represents paid participation in off-farm work, and 'zero' represents no paid off-farm work.
2) Model c) is re-estimated, with a trichotomous dependent variable (on-farm paid, on-farm unpaid and off-farm work) and results are compared with those from the first step.
3) Model c) is estimated once more with a polychotomous dependent variable (six categories: on-farm paid, on farm unpaid, farm wage earner, non-agricultural wage earner, self-employed sole trader and self-employed employer). It is expected that the increasing division of options should lead to sharper and more interesting results.

The coefficients in all of these estimated reduced form participation models may be interpreted as follows: for a dichotomous *participation model* a variable will have a *negative* sign if any one or more of the following effects is dominant (i.e. is greater than the sum of any countervailing effects) – i) the variable's effect is to decrease rewards for given levels of off-farm work (e.g. being young), ii) it increases farm profits (e.g. having a large, well-capitalised farm) and iii) it increases the shadow wage (e.g. greater unearned income). Some variables (e.g. education) will probably

[16] Equation 1-12 (a household level multinomial logit) is not used in any estimation.

operate through all three channels. Another possibility, iv), is that there is some form of rationing of off-farm work in place according to the variable in question (e.g. being from a particular group might increase off-farm rewards, but people from that group may be rationed out of the categories most profitable for them, and so a group dummy may end up having a negative sign in a participation model). This possibility is examined in Chapter 5.

In any one case, interpretation of the channels through which variables affect participation may be problematic, since all the reduced form tells us is the overall effect; nevertheless, this overall effect may be interesting in itself, and in many cases inferences based on background knowledge and knowledge of stylised facts from the literature may help to fine tune interpretations of results.

This particular type of disaggregation – the disaggregation of choices – is not normally carried out within the farm household tradition though, as mentioned in the introduction to this chapter, it does tend to be carried out within the rural livelihoods literature (and is recommended by Reardon et. al., 2001). Adapting this type of disaggregation to participation models in the farm household model is straightforward. Instead of comparing wages and reservation wages (as in Equations 1-5 and 1-10), one ought to compare returns in each of the disaggregated options along with reservation wages. The assumption of the typical random utility model is that the agent enters the category with the highest utility. In the case of some categories, however, for some people, it could be they are forced by custom or family pressure into a category (e.g. unpaid farm worker) or they are prevented by transaction costs, or by custom or tradition, from entering a category. In the reduced form setting of this chapter it is not possible to deal with these questions. Some of them are broached in the next chapter but for now conventional random utility interpretations suffice.

4.2 Participation estimation results

In this section the results of the participation models are reported in some detail, unfortunately with some unavoidable repetition. In the concluding discussion of the chapter all the results are brought together.

4.2.1 Model a) Household participation in off-farm work:

In all the models without random effects in this chapter, errors are assumed to be independent between segmentos, where the segmento is the primary sampling unit (a small rural community usually), but not within them. Also, many of the models have two versions – a main version, and one that additionally tests the influence of two variables that might be indicative of the balance of power in the household – the local male/female ratio and whether a person has married into the area. The results from these additional tests are not very clear but are reported anyway because some of the ancillary information about people who marry in is useful.

Table 4-1 below shows the results of the household participation probit models, with and without individual 'segmento' level random effects. Also included, for illustrative purposes, are versions of the same two models where three potentially endogenous variables – the degree of crop diversification, the proportion of crops sold and the level of farm capital – are instrumented (by variables listed in footnote 17).

Table 4-1: Household Participation Model

	Probit	Random Effects Probit	IV model – Probit	IV model – Random Effects Probit
Age	0.061** [0.01]	0.068*** [0.01]	0.046* [0.07]	0.049* [0.07]
Age^2	-0.001*** [0.01]	-0.001*** [0.00]	-0.001** [0.03]	-0.001** [0.03]
Gender (1=male)	0.552** [0.01]	0.636** [0.01]	0.509** [0.04]	0.565** [0.03]
Years of education	0.067*** [0.00]	0.076*** [0.00]	0.059*** [0.01]	0.066** [0.01]
Years experience on-farm	-0.022*** [0.00]	-0.025*** [0.00]	-0.022*** [0.00]	-0.023*** [0.00]
Number of adult family members (15-65)	0.059* [0.10]	0.075* [0.06]	0.053 [0.22]	0.063 [0.18]
Number of children 15 years old or less	0.016 [0.53]	0.017 [0.58]	0.026 [0.39]	0.028 [0.38]
Spanish speaking	0.002 [0.99]	-0.066 [0.64]	-0.073 [0.81]	-0.202 [0.59]
Unearned income	-0.000* [0.08]	0.000 [0.17]	0.000 [0.11]	0.000 [0.15]
Home owners	-0.253 [0.12]	-0.242 [0.18]	-0.273 [0.16]	-0.219 [0.31]
Number of rooms in the family home	0.061* [0.09]	0.057 [0.16]	0.036 [0.43]	0.035 [0.49]
Farm capital	-0.001*** [0.00]	-0.001*** [0.00]	0.0001 [0.82]	0.001 [0.97]
Farm size	-0.013 [0.10]	-0.010 [0.20]	-0.021 [0.14]	-0.019 [0.29]
Change in farm size (1991-1994)	-0.019 [0.44]	-0.016 [0.63]	-0.004 [0.91]	0.001 [0.99]
Proportion of farm output sold on the market	-0.606*** [0.00]	-0.619*** [0.00]	-0.442 [0.66]	-0.109 [0.93]
Inverted Simpson Index	-0.07 [0.11]	-0.087* [0.09]	-0.213 [0.33]	-0.187 [0.50]
Coastal region	0.049 [0.80]	0.082 [0.70]	-0.107 [0.74]	-0.170 [0.68]
Rainforest region	0.228 [0.21]	0.296 [0.16]	0.266 [0.36]	0.299 [0.42]
Importance of local non-agricultural income sources	0.502 [0.18]	0.641 [0.14]	0.156 [0.80]	0.231 [0.77]
Difficulty of access (distance)	0.077 [0.17]	0.102 [0.14]	0.085 [0.12]	0.103 [0.14]

	Probit	Random Effects Probit	IV model – Probit	IV model – Random Effects Probit
Difficulty of access (time)	-0.032** [0.04]	-0.031 [0.23]	-0.025 [0.30]	-0.03 [0.31]
Farm tenure (higher means less secure)	0.082** [0.01]	0.068* [0.07]	0.090** [0.02]	0.075* [0.07]
Maize	0.088 [0.48]	0.059 [0.69]	0.015 [0.92]	0.013 [0.94]
Potato	0.039 [0.78]	0.123 [0.43]	-0.012 [0.94]	0.078 [0.66]
Rice	-0.365** [0.05]	-0.372* [0.05]	-0.449** [0.02]	-0.409* [0.07]
Banana	0.494 [0.18]	0.439 [0.13]	0.623* [0.06]	0.598 [0.11]
Intercept	-1.402* [0.06]	-1.758** [0.02]	-0.424 [0.79]	-0.840 [0.67]
N	980	980	970	970
Number of SEG		109		109

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported.

In the first probit model, we find that a household is more likely to participate in off-farm work if the following conditions hold: there are more adult males than females in the household, and more adults overall; household members have more education on average and less on-farm experience, as well as less unearned income; the farm is not highly capitalized; only a small proportion of crops is sold; the farm is not in a remote area as far as public transport is concerned; the household has a large house but weak farm tenure and the main crop is not rice. The typical concave age pattern also holds. The overall sense from these preliminary results is of farming as an alternative to, rather than complement of, off-farm work. All else equal, small farmers, with little capital and weak tenure, tend to work off farm.

These results are altered slightly when we look at the random effects estimation, but the overall picture is not seriously altered. Three variables become insignificant – unearned income, the size of the house, and the public transport variable measuring the degree of remoteness. One variable becomes significant – the inverted Simpson index (the degree of diversification). The standard test statistics suggest that the segmento level random effects are significant, and so the random effects model is preferable.

When we turn to instrumental variable versions of the two models, we find that the three instrumented farm level variables (farm capital, proportion of crops sold and the diversification index) are no longer significant.[17] This is either because the

[17] The first step regressions for the IV models all had adjusted r^2s between .15 and .2, and are reported in the appendix to this chapter. The variables used in all three regressions were the X variables already presented as well as five new segmento level variables – average distance from a variety of facilities, men's agricultural wage, women's agricultural wage, the number of dwellings in the segmento and the number of times a week public transport comes to the nearest pick-up point - and a farm level variable roughly indicating farm efficiency and/or land quality (net output per hectare), as well as total members of the family, whether resident or not. Reported standard errors are corrected for the two step procedure in the standard probit model, but not the random effects probit.

instruments are weak or because the variables indeed are wholly endogenous. The former is much more likely. Given the way land was distributed in the land reforms (as noted in Chapter 2), complete endogeneity seems extremely unlikely for the diversification variable (communities and individuals received land at different altitudes). For the farm capital and market embeddedness variables exogeneity is also probable. Farm capital is higher for full time farm households than others (Table 2-6), so off-farm work is unlikely to be a major determinant of the level farm capital. The degree of market embeddedness is associated, mainly, with region and language spoken, both of which are also largely exogenous. Despite some (very) slight doubts, these three variables are left in all the other versions of the model because often it is precisely the interplay between off-farm participation choices and these variables that is of special interest (and, given the size of the model, it is unlikely that the coefficients estimated for the other variables will be seriously biased: in Table 4-1 in the IV models, all variables significant at 5%, apart from the three instrumented ones and the remoteness variable, are significant in all equations, with coefficients that vary little across equations).

4.2.2 Model b) The main couple's participation decisions

In the Bivariate Probit models reported in Table 4-2, main male and female (operator and spouse) participation is jointly modelled.

Table 4-2: Participation of Operator and Spouse

Probit models of participation		Univariate		Bivariate (without bargaining variables)		Bivariate (with bargaining variables)	
		Operator	Spouse	Operator	Spouse	Operator	Spouse
Age	Operator	0.016 [0.72]	0.089** [0.04]	0.02 [0.65]	0.091** [0.03]	0.014 [0.76]	0.095** [0.03]
Age^2	Operator	0.001 [0.50]	-0.001** [0.02]	0.001 [0.42]	-0.001** [0.02]	0.001 [0.52]	-0.001** [0.02]
Age	Spouse	0.008 [0.85]	0.004 [0.93]	0.013 [0.76]	0.002 [0.97]	0.019 [0.66]	-0.001 [0.98]
Age^2	Spouse	0.001 [0.91]	0.001 [0.72]	0.001 [0.84]	0.001 [0.71]	0.001 [0.73]	0.001 [0.76]
Years of education	Operator	0.012 [0.53]	-0.021 [0.29]	0.006 [0.77]	-0.017 [0.39]	0.008 [0.70]	-0.011 [0.59]
Years of education	Spouse	0.008 [0.65]	0.039** [0.03]	0.011 [0.56]	0.036** [0.04]	0.012 [0.52]	0.036** [0.05]
Years experience on-farm	Operator	-0.018*** [0.00]	0.019*** [0.00]	-0.018*** [0.00]	0.018*** [0.00]	-0.018*** [0.00]	0.019*** [0.00]
Years experience on-farm	Spouse	-0.001 [0.80]	-0.052*** [0.00]	0.000 [0.96]	-0.053*** [0.00]	0.000 [0.98]	-0.055*** [0.00]
Number of people over 65 in the house		0.054 [0.75]	0.163 [0.40]	0.090 [0.60]	0.197 [0.32]	0.099 [0.57]	0.192 [0.35]
Number of children (under 15)		-0.023 [0.71]	-0.007 [0.90]	-0.022 [0.71]	-0.005 [0.93]	-0.021 [0.73]	0.014 [0.80]
Number of children (under 7)		0.064 [0.47]	0.004 [0.96]	0.065 [0.45]	0.007 [0.93]	0.056 [0.52]	-0.015 [0.86]
Unearned income		0.001 [0.28]	0.001 [0.58]	0.001 [0.30]	0.001 [0.61]	0.001 [0.28]	0.001 [0.55]
Number of rooms in the family home		0.006 [0.94]	-0.001 [1.00]	0.003 [0.98]	-0.013 [0.90]	0.008 [0.94]	-0.068 [0.51]

Probit models of participation		Univariate		Bivariate (without bargaining variables)		Bivariate (with bargaining variables)	
		Operator	Spouse	Operator	Spouse	Operator	Spouse
Spanish speaking		0.358*** [0.01]	0.274* [0.06]	0.370*** [0.01]	0.271* [0.07]	0.386*** [0.01]	0.287* [0.06]
Home owners		-0.096 [0.60]	-0.230 [0.25]	-0.102 [0.59]	-0.249 [0.20]	-0.092 [0.64]	-0.288 [0.11]
Farm capital		-0.000** [0.03]	0.000 [0.12]	-0.000** [0.03]	-0.000* [0.10]	-0.000** [0.04]	0.000 [0.13]
Farm size		-0.041** [0.01]	-0.030* [0.08]	-0.040*** [0.01]	-0.030* [0.06]	-0.042*** [0.00]	-0.028* [0.09]
Inverted Simpson Index		-0.026 [0.64]	-0.030 [0.64]	-0.021 [0.72]	-0.033 [0.61]	-0.013 [0.82]	-0.041 [0.52]
Proportion of farm output sold on the market		-1.251*** [0.00]	-0.686*** [0.01]	-1.247*** [0.00]	-0.675*** [0.01]	-1.279*** [0.00]	-0.733*** [0.01]
Rice		-0.406* [0.06]	-0.302 [0.20]	-0.410* [0.05]	-0.303 [0.19]	-0.431** [0.04]	-0.326 [0.15]
Maize		0.266 [0.15]	-0.271 [0.13]	0.244 [0.19]	-0.292 [0.11]	0.232 [0.21]	-0.287 [0.11]
Banana		0.400 [0.27]	0.735** [0.02]	0.428 [0.23]	0.732** [0.02]	0.421 [0.25]	0.772** [0.02]
Potato		-0.004 [0.98]	-0.184 [0.29]	-0.004 [0.98]	-0.223 [0.21]	-0.007 [0.97]	-0.217 [0.23]
Importance of local non-agricultural income sources		1.103*** [0.00]	0.876** [0.01]	1.111*** [0.00]	0.896** [0.01]	1.149*** [0.00]	0.721** [0.05]
Difficulty of access (distance)		-0.072 [0.25]	-0.067 [0.23]	-0.061 [0.32]	-0.061 [0.27]	-0.061 [0.36]	-0.110* [0.07]
Difficulty of access (time)		-0.003 [0.86]	0.004 [0.86]	-0.004 [0.85]	0.003 [0.91]	-0.003 [0.90]	0.001 [0.99]
Farm tenure (higher means less secure)		0.006 [0.89]	0.013 [0.75]	0.003 [0.95]	0.015 [0.72]	0.003 [0.94]	0.013 [0.74]
Coastal region		0.338 [0.16]	-0.048 [0.87]	0.343 [0.15]	-0.098 [0.75]	0.333 [0.17]	0.013 [0.97]
Rainforest region		0.270 [0.20]	-0.087 [0.69]	0.267 [0.21]	-0.134 [0.53]	0.202 [0.36]	-0.076 [0.74]
Unpaid labour exchange is customary		0.172 [0.26]	-0.273 [0.11]	0.156 [0.31]	-0.284* [0.09]	0.183 [0.25]	-0.272 [0.11]
Gender ratio in village (male/female)						-0.136 [0.46]	-0.599*** [0.01]
Migrated into area to marry	Operator					0.214 [0.12]	-0.105 [0.53]
	Spouse					-0.115 [0.73]	0.305 [0.36]
Change in farm size (1991-1994)						0.025 [0.64]	-0.055 [0.14]
Intercept		-0.103 [0.89]	-2.046** [0.03]	-0.270 [0.72]	-1.991** [0.03]	-0.270 [0.74]	-1.073 [0.27]
N		575	575	575	575	575	575

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported. There is a positive correlation coefficient of .42 and .425, respectively, in the two bivariate models, significant at 1%.

In the initial model, the operator is more likely to participate in off-farm work if he is less experienced on the farm. He is also more likely to participate if he lives in a large house, has little farm capital, is on a smaller farm (farm size was not significant at household level), does not sell what crops he produces; all else equal, he works in an area where farming is less important and his farm will not produce rice.

In turn, for the spouse to work off-farm, she is likely to be living with an experienced on-farm operator but be inexperienced herself; she will have little farm capital on a small farm, and the farm will produce less crops for the market; she will live in an area where farming is not so important. If the main crop is bananas, she is more likely to work off-farm. Overall, her decisions appear to be influenced by more or less the same variables as the operator, as well as by his age and experience. The (unreported) correlation variable (rho) is highly significant, suggesting that spousal decisions are inter-related.

When we add in the two variables that could be indicative of a Collective model (ratio of males to females in the local community, and operator or spouse marrying into the area), we find evidence the first variable significant for spouses – they are less likely to work off farm if there are more men than women in the area.

Three differences between the results reported in this table and those reported in the previous one are the effects of education, experience and language. The education effect, from Table 4-1 does appear to be quite strong for the household as a whole, but in Table 4-2 we find that this effect is not strong for the operator but only for the spouse. This effect is not due to the interaction term as separate probits also reveal that education is insignificant for the operator. The topic will be returned to in the analysis of later models. The experience of the operator seems to determine both his own and the spouse's time participation decision, suggesting that – apart from any simultaneity in decision making – intra-household comparative advantage probably affects participation decisions. Finally, the language variable is significant at individual level, but not at household level. Since it only takes one member of a couple to participate to ensure participation of the household, the individual regressions should provide a more reliable picture.

4.2.3 Model c) Individuals within the household

In Tables 4-3 to 4-7, participation for all working individuals in the family is modelled in three steps – firstly as a dichotomous variable (Tables 4.3 and 4.4), secondly as a trichotomous variable (Tables 4.5 and 4.6), and thirdly as a polychotomous variable (Tables 4.7). In these models it should be noted that, apart from intercepts, coefficients are restricted to be the same for all family members (an extra year of education, for example, is restricted to having the same coefficient for the operator, spouse and adult child), and intra-household interdependencies are modelled in two ways: as household random effects in Tables 4.4 and 4.6 and as non-independent errors *within* households in Tables 4.3, 4.5 and 4.7. The sample is different from the bivariate model, because it also includes operators from households with no spouses, spouses from households with no operators, and also includes other adults (over 14s), the vast majority of whom are adult children of the main couple.

These models also include a few variables not included previously (the average education of others in the house, average on-farm experience of others in the house and an 'at school' dummy variable for young adults – 15-18 year olds). Quantitative household and farm variables are included as 'averages per working adult' (e.g. average farm capital per working adult, or number of rooms per adult), although all the models have also been estimated using household or farm totals, controlling for number of adults, rather than per adult averages, with little difference in results.

4.2.3.1 *Model c) Dichotomous dependent variables*

From Tables 4.3 and 4.4, we find that results are strongly affected by the inclusion or otherwise of unpaid farm workers in the zero category of these dichotomous models (where 'one' means engaged in off-farm work).

Table 4-3: Participation of Individuals

Dependent Variable = participation	Only Paid People in sample	Paid+Unpaid	Only Paid (+'bargaining variables')	Paid+Unpaid (+'bargaining variables')
Age	0.034 [0.19]	0.119*** [0.00]	0.034 [0.19]	0.121*** [0.00]
Age^2	0.000 [0.10]	-0.001*** [0.00]	-0.000* [0.10]	-0.001*** [0.00]
Attending school	1.529** [0.03]	-0.723*** [0.00]	1.509** [0.03]	-0.719*** [0.00]
Spouse	0.338* [0.05]	-1.088*** [0.00]	0.043 [0.88]	-1.049*** [0.00]
Other male (mainly adult children)	0.595** [0.02]	-0.747*** [0.00]	0.390 [0.17]	-0.782*** [0.00]
Other female (mainly adult children)	1.425*** [0.00]	-0.780*** [0.00]	1.219*** [0.00]	-0.814*** [0.00]
Years of education	0.027 [0.15]	0.037*** [0.00]	0.026 [0.16]	0.037*** [0.01]
Years experience on-farm	-0.036*** [0.00]	-0.050*** [0.00]	-0.035*** [0.00]	-0.050*** [0.00]
Average years of education of rest of household	0.022 [0.28]	-0.004 [0.82]	0.023 [0.26]	-0.002 [0.90]
Average on-farm experience of rest of household	0.007 [0.25]	-0.007 [0.19]	0.008 [0.21]	-0.006 [0.21]
Number of people over 65 in the house	0.114 [0.56]	-0.054 [0.69]	0.124 [0.52]	-0.055 [0.68]
Number of children (under 15)	-0.018 [0.77]	-0.018 [0.70]	-0.013 [0.82]	-0.015 [0.74]
Number of children (under 7)	0.054 [0.54]	0.040 [0.57]	0.050 [0.57]	0.041 [0.56]
Unearned income	-0.001*** [0.01]	-0.001* [0.07]	-0.001*** [0.01]	-0.001* [0.06]
Number of rooms in the family home	0.083 [0.26]	0.199*** [0.00]	0.085 [0.25]	0.198*** [0.00]
Spanish speaking	0.088 [0.61]	0.266* [0.05]	0.089 [0.61]	0.259* [0.06]
Home owners	-0.267 [0.21]	-0.273 [0.12]	-0.261 [0.22]	-0.259 [0.14]
Farm capital per adult	-0.000** [0.02]	-0.000*** [0.01]	-0.000** [0.03]	-0.000** [0.01]
Farm size per adult	-0.053** [0.03]	-0.045** [0.02]	-0.054** [0.02]	-0.046** [0.02]
Change in farm size (1991-1994)	-0.016 [0.70]	-0.029 [0.38]	-0.013 [0.75]	-0.028 [0.39]
Inverted Simpson Index	-0.212*** [0.00]	-0.154*** [0.00]	-0.212*** [0.00]	-0.155*** [0.00]
Proportion of farm output sold on the market	-1.281*** [0.00]	-0.867*** [0.00]	-1.304*** [0.00]	-0.873*** [0.00]
Rice	-0.690*** [0.00]	-0.699*** [0.00]	-0.708*** [0.00]	-0.712*** [0.00]

Dependent Variable = participation	Only Paid People in sample	Paid+Unpaid	Only Paid (+'bargaining variables')	Paid+Unpaid (+'bargaining variables')
Maize	0.399** [0.04]	0.156 [0.31]	0.394** [0.04]	0.154 [0.32]
Banana	-0.196 [0.52]	-0.141 [0.56]	-0.199 [0.52]	-0.142 [0.56]
Potato	0.133 [0.50]	-0.047 [0.78]	0.135 [0.50]	-0.045 [0.79]
Importance of local non-agricultural income sources	1.175*** [0.01]	0.863*** [0.00]	1.119** [0.01]	0.817*** [0.01]
Difficulty of access (distance)	-0.036 [0.62]	-0.005 [0.93]	-0.057 [0.44]	-0.016 [0.79]
Difficulty of access (time)	-0.037 [0.10]	-0.008 [0.67]	-0.038* [0.10]	-0.006 [0.74]
Farm tenure (higher means less secure)	-0.002 [0.97]	0.046 [0.24]	-0.003 [0.94]	0.043 [0.27]
Coastal region	0.088 [0.71]	-0.087 [0.67]	0.165 [0.49]	-0.044 [0.84]
Rainforest region	0.294 [0.18]	0.009 [0.96]	0.333 [0.13]	0.032 [0.87]
Unpaid labour exchange is customary	0.172 [0.34]	-0.17 [0.24]	0.18 [0.32]	-0.169 [0.24]
No. of men relative to women in area - men			-0.449* [0.09]	-0.104 [0.66]
No. of men relative to women in area - women			0.179 [0.61]	-0.241 [0.34]
Migrated into area to get married			0.128 [0.75]	0.661* [0.10]
Intercept	0.485 [0.56]	-1.421** [0.02]	0.903 [0.32]	-1.194* [0.08]
N	1538	2344	1538	2344

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported. The excluded group is those who farm for returns, and do not work for wages off-farm.

If the sample is restricted to those reporting themselves financially rewarded for their work (and un paid farm work is treated as another form of housework, and excluded), then being the main female or 'other' female in a household leads to a greater probability of working off the farm (than being the main male, or 'operator'). Other males also are more likely to work off farm. On-farm experience has the familiar negative effect. Neither the education nor experience of others in the house appears to have any effect. Unexpected unearned income here has the expected negative effect (for the first time since the non-random effects models of Table 4-1). Farm capital, too, has a negative sign, as do diversification and proportion of crops sold. Farm size is also clearly significant, negative. The importance of farming in the area has the expected sign. Rice farming again has a negative effect; maize farming positive. When we include the intra-family 'bargaining' variables in the model, we find that men work less off-farm if there are more men in the area.

When unpaid farm work is subsumed within the zero category (unpaid farm work is treated, that is, as being in the same category as paid farm work) the main effect is on the results for the intra-family dummies. Everyone else on the farm is far less likely to work off-farm than the main male. Education becomes significant as does the Spanish speaking dummy and the house size variable (number of rooms) while the maize variable coefficients become larger and significant (suggesting, if the specification is good, that, compared to other farms, in farms where unpaid adult family farm work is common it is usual to have adults who are more experienced in farming and adults who have less education and wealth, grow more maize and do not speak Spanish at home). The main farm level variables (size, capital level, diversification and proportion of crops sold) do not differ in significance across samples, suggesting that these do not differ greatly according to whether unpaid work is common or not. Rice farming once again has a considerable negative effect.

When the family 'bargaining' variables are included, we find that marrying into the area appears to have a positive effect on participation. The positive sign on the marrying-in variable when unpaid farm work is in the sample, and its insignificance when unpaid farm work is not in the sample, suggest that this variable probably does not have an intra-family 'bargaining' interpretation, unless we accept that marrying in will increase bargaining power. It is more likely that the type of people who marry in are unlikely to also be the type who will do not unpaid farm work, but are likely to work off-farm.

The main differences with the probit models for couples only are in the effects of house size, unearned income and crop diversification. That these were not significant before is partly due to there being a larger sample of both operators and spouses in the latest regressions (only houses with couples who both did farm work are used in the bivariate probits), and because adult children are now included. As in the move from the household level equation to the couple equation, extending the sample brings sharper results.

Table 4-4: Random effects individual participation models

Dependent Variable: Participation	Only Paid People in sample	Paid+Unpaid	Only Paid (+'bargaining variables')	Paid+Unpaid (+'bargaining variables')
Age	0.036 [0.22]	0.138*** [0.00]	0.037 [0.21]	0.140*** [0.00]
Age^2	0.000 [0.12]	-0.001*** [0.00]	0.000 [0.11]	-0.001*** [0.00]
Attending school	1.575* [0.06]	-0.806*** [0.00]	1.553* [0.06]	-0.804*** [0.00]
Spouse	0.377* [0.06]	-1.280*** [0.00]	0.043 [0.89]	-1.233*** [0.00]
Other male	0.608** [0.02]	-0.971*** [0.00]	0.392 [0.19]	-0.994*** [0.00]
Other female	1.467*** [0.00]	-1.044*** [0.00]	1.251*** [0.00]	-1.065*** [0.00]
Years of education	0.034 [0.11]	0.046*** [0.00]	0.033 [0.12]	0.045*** [0.00]
Years experience on-farm	-0.040*** [0.00]	-0.059*** [0.00]	-0.040*** [0.00]	-0.059*** [0.00]
Average years of education of rest of household	0.021 [0.35]	-0.003 [0.86]	0.022 [0.33]	-0.002 [0.93]
Average on-farm experience of rest of household	0.008 [0.22]	-0.005 [0.40]	0.009 [0.18]	-0.004 [0.43]

Dependent Variable: Participation	Only Paid People in sample	Paid+Unpaid	Only Paid (+'bargaining variables')	Paid+Unpaid (+'bargaining variables')
Number of people over 65 in the house	0.106 [0.58]	-0.082 [0.58]	0.119 [0.54]	-0.084 [0.57]
Number of children (under 15)	-0.018 [0.79]	-0.028 [0.62]	-0.013 [0.85]	-0.024 [0.67]
Number of children (under 7)	0.061 [0.54]	0.052 [0.53]	0.056 [0.57]	0.053 [0.52]
Unearned income	-0.001*** [0.01]	-0.001** [0.05]	-0.001*** [0.01]	-0.001** [0.04]
Number of rooms in the family home per adult	0.091 [0.26]	0.233*** [0.00]	0.092 [0.26]	0.230*** [0.00]
Spanish speaking	0.085 [0.64]	0.307* [0.05]	0.087 [0.64]	0.297* [0.06]
Home owners	-0.297 [0.20]	-0.347* [0.09]	-0.29 [0.21]	-0.328 [0.11]
Farm capital per adult	-0.001*** [0.00]	-0.001*** [0.00]	-0.001*** [0.00]	-0.001*** [0.00]
Farm size per adult	-0.055*** [0.00]	-0.049*** [0.00]	-0.057*** [0.00]	-0.050*** [0.00]
Change in farm size (1991-1994)	-0.020 [0.64]	-0.039 [0.37]	-0.018 [0.69]	-0.039 [0.38]
Inverted Simpson Index	-0.234*** [0.00]	-0.180*** [0.00]	-0.234*** [0.00]	-0.182*** [0.00]
Proportion of farm output sold on the market	-1.378*** [0.00]	-0.992*** [0.00]	-1.405*** [0.00]	-1.001*** [0.00]
Rice	-0.758*** [0.00]	-0.789*** [0.00]	-0.775*** [0.00]	-0.807*** [0.00]
Maize	0.444** [0.04]	0.178 [0.31]	0.439** [0.04]	0.177 [0.32]
Banana	-0.223 [0.51]	-0.201 [0.50]	-0.228 [0.50]	-0.201 [0.50]
Potato	0.138 [0.53]	-0.082 [0.65]	0.139 [0.53]	-0.080 [0.66]
Importance of local non-agricultural income sources	1.313*** [0.01]	1.110*** [0.00]	1.254*** [0.01]	1.060*** [0.01]
Difficulty of access (distance)	-0.043 [0.58]	-0.001 [0.99]	-0.065 [0.41]	-0.012 [0.86]
Difficulty of access (time)	-0.041 [0.15]	-0.016 [0.49]	-0.041 [0.14]	-0.014 [0.57]
Farm tenure (higher means less secure)	0.002 [0.96]	0.056 [0.18]	0.001 [0.99]	0.053 [0.21]
Coastal region	0.071 [0.78]	-0.136 [0.56]	0.151 [0.57]	-0.086 [0.72]
Rainforest region	0.320 [0.19]	0.030 [0.89]	0.361 [0.15]	0.055 [0.80]
Unpaid labour exchange is customary	0.177 [0.36]	-0.202 [0.23]	0.186 [0.33]	-0.201 [0.24]
No. of men relative to women in area - men			-0.475* [0.10]	-0.084 [0.77]
No. of men relative to women in area - women			0.230 [0.57]	-0.262 [0.40]
Migrated into area to get married			0.206 [0.66]	0.892** [0.04]
Intercept	0.520 [0.58]	-1.765** [0.02]	0.939 [0.35]	-1.556* [0.06]
N	1538	2344	1538	2344
Number of HID	993	1036	993	1036

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported. The excluded group is those who farm for returns, and do not work for wages off-farm.

The standard LR tests suggest that all four random effects models reported in Table 4-4 are improvements on the four equivalent models reported in Table 4-3 (the p-values on the null of no random effects are both .05 for the smaller models and zero for the larger ones). However, the size of the coefficients and the list of significant variables are not very different when we compare the two tables. Comparison of the two tables reveals no important new information from the household random effects specification.

4.2.3.2 *Model c) Trichotomous dependent variables*

The sensitivity of the results in Tables 4.3 and 4.4 to whether unpaid farm work is included or not suggests a trichotomous approach could be profitable. In Tables 4.5 and 4.6 the dependent variable has three values – paid on-farm work only, unpaid on-farm work only and paid off-farm work. A multinomial logit model is used for the estimations, without household level random effects in Table 4-5 and with, in Table 4-6. The first category – paid on-farm work – is the base category, so the coefficients are set at zero for this value of the dependent variable. Hausman's test for the independence of irrelevant alternatives is carried out for all the multinomial models and is never rejected.

Table 4-5: Multinomial logit individual participation (3 categories)

Dependent variable	Unpaid farm work	Off-farm work	Unpaid farm work (+bargaining variables')	Off-farm work (+'bargaining variables')
Age	-0.092*** [0.00]	0.048* [0.08]	-0.096*** [0.00]	0.049* [0.07]
Age^2	0.001* [0.07]	-0.001** [0.04]	0.001* [0.05]	-0.001** [0.03]
Attending school	1.976*** [0.01]	1.289* [0.08]	1.954*** [0.01]	1.286* [0.08]
Spouse	3.881*** [0.00]	0.244 [0.17]	2.769*** [0.00]	-0.082 [0.78]
Other male	3.963*** [0.00]	0.684** [0.01]	3.178*** [0.00]	0.476 [0.12]
Other female	4.686*** [0.00]	1.363*** [0.00]	3.905*** [0.00]	1.157*** [0.01]
Years of education	-0.021 [0.31]	0.022 [0.25]	-0.022 [0.30]	0.02 [0.29]
Years experience on-farm	0.017*** [0.01]	-0.042*** [0.00]	0.018*** [0.01]	-0.042*** [0.00]
Average years of education of rest of household	0.044* [0.10]	0.017 [0.42]	0.044 [0.10]	0.017 [0.41]
Average on-farm experience of rest of household	0.020*** [0.00]	0.004 [0.52]	0.020*** [0.00]	0.004 [0.50]
Number of people over 65 in the house	0.27 [0.19]	0.123 [0.54]	0.293 [0.15]	0.134 [0.49]
Number of children (under 15)	0.007 [0.92]	-0.003 [0.95]	-0.001 [0.99]	0.000 [1.00]
Number of children (under 7)	0.013 [0.90]	0.028 [0.74]	0.017 [0.86]	0.023 [0.78]
Unearned income	-0.001* [0.07]	-0.001*** [0.01]	-0.001* [0.06]	-0.001*** [0.01]
Number of rooms in the family home per adult	-0.387*** [0.00]	0.060 [0.42]	-0.373*** [0.00]	0.063 [0.40]
Spanish speaking	-0.361* [0.08]	0.067 [0.71]	-0.376* [0.07]	0.069 [0.70]
Home owners	-0.06 [0.83]	-0.281 [0.19]	-0.076 [0.78]	-0.276 [0.20]
Farm capital per adult	0.000 [0.81]	-0.001** [0.02]	0.000 [0.90]	-0.001** [0.03]
Farm size per adult	-0.023 [0.31]	-0.049** [0.03]	-0.023 [0.28]	-0.051** [0.02]
Change in farm size (1991-1994)	0.046 [0.30]	-0.020 [0.62]	0.050 [0.25]	-0.017 [0.67]
Inverted Simpson Index	-0.136* [0.08]	-0.234*** [0.00]	-0.142* [0.07]	-0.234*** [0.00]
Proportion of farm output sold on the market	-0.852*** [0.00]	-1.380*** [0.00]	-0.856*** [0.00]	-1.409*** [0.00]
Rice	-0.253 [0.32]	-0.780*** [0.00]	-0.233 [0.36]	-0.787*** [0.00]
Maize	0.273 [0.21]	0.342* [0.08]	0.277 [0.20]	0.338* [0.08]
Banana	-0.164	-0.223	-0.168	-0.228

Dependent variable	Unpaid farm work	Off-farm work	Unpaid farm work (+bargaining variables')	Off-farm work (+'bargaining variables')
	[0.64]	[0.47]	[0.63]	[0.46]
Potato	0.296 [0.20]	0.177 [0.38]	0.314 [0.18]	0.192 [0.34]
Importance of local non-agricultural income sources	0.455 [0.30]	1.091** [0.01]	0.429 [0.33]	1.046** [0.02]
Difficulty of access (distance)	-0.057 [0.51]	-0.035 [0.63]	-0.059 [0.51]	-0.055 [0.46]
Difficulty of access (time)	-0.052* [0.06]	-0.032 [0.15]	-0.051* [0.06]	-0.033 [0.14]
Farm tenure (higher means less secure)	-0.093* [0.07]	-0.001 [0.98]	-0.091* [0.07]	-0.002 [0.96]
Coastal region	0.259 [0.35]	0.106 [0.66]	0.306 [0.26]	0.174 [0.48]
Rainforest region	0.444* [0.09]	0.233 [0.30]	0.471* [0.07]	0.271 [0.23]
Unpaid labour exchange is customary	0.444** [0.03]	0.090 [0.62]	0.437** [0.04]	0.096 [0.60]
No. of men relative to women in area - men			-1.766** [0.01]	-0.449* [0.10]
No. of men relative to women in area - women			0.854** [0.05]	0.271 [0.50]
Migrated into area to get married			-1.118** 0.038	0.127 0.755
Intercept	0.157 0.873	0.749 0.378	1.529 [0.19]	1.117 [0.23]
N	2344	2344	2344	2344

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported. The excluded group is those who farm for returns, and do not work for wages off-farm.

In examining the various results of Table 4-5, one can see that the results of the previous two tables are cast in clearer relief. The effect, for instance, of being a spouse or adult child in the household is to clearly increase the likelihood of being in the unpaid on-farm worker category. Length of on-farm experience has a positive effect on the likelihood of doing unpaid work on the farm. Non-Spanish speakers in smaller houses, who farm non-commercially, are also more likely to enter this category. For the first time the education and on-farm experience of others in the house are significant. One is more likely to work unpaid if others have more education or experience.

Own completed education does not appear to affect entry into the unpaid category (compared to the base category), although being at school, not surprisingly, increases the relative probability of entering it. Unexpected unearned income, as one might expect, decreases the chances of entering this category, as does having a larger house. None of the crop variables are significant but the rainforest variable is significant positive.

In areas with more males than females, (female) spouses are more likely to enter the unpaid category than to be paid farmers; but they are approximately equally likely to work off-farm. Males are more likely to work on the farm for reward in these areas. Those who marry in are not likely to enter the unpaid category. This would appear to suggest that these two variables – particularly, again, the latter – may not carry an intra-family 'bargaining' interpretation. Finally, it is noticeable with regard to the unpaid category that the familiar positive-negative signs on the coefficients of *age* and *age squared* are reversed, suggesting people become more and more likely to enter this category as they get older.

With regard to work off the farm, unsurprisingly the results are similar to those for the smaller sample in Table 4-3 and Table 4-4. To summarize briefly: it would appear that adult children are more likely to work off-farm than parents. Unexpected 'unearned' income reduces participation. Speaking Spanish appears to have no effect (because unpaid work is an option). Once again, being from a well-capitalised, large farm reduces chances of working off-farm. Being from a commercial farm has the familiar negative 'effect' on off-farm participation (stronger than on unpaid family farm work) as does the crop diversification variable. Rice farming reduces entry into off-farm work, while maize farming increases it. The presence of 'minka' or a local culture of informal labour exchange is insignificant for off-farm work, although it increases the probability of doing unpaid work.

Table 4-6: random effects multinomial logit (3 categories)

	Random Effects Multinomial Logit	
Dependent Variable: participation	**Unpaid farm work**	**Off-farm work**
Age	-0.127*** [0.00]	0.045 [0.13]
Age^2	0.001*** [0.00]	-0.001 [0.09]
Attending school	1.754** [0.02]	1.166 [0.13]
Spouse	3.027*** [0.00]	-1.156*** [0.01]
Other male	3.846*** [0.00]	0.262 [0.36]
Other female	4.583*** [0.00]	0.878** [0.03]
Years of education (> 8)	-0.093 [0.60]	0.213 [0.17]
Years of education (> 13)	0.044 [0.88]	0.303 [0.26]
Years of education (> 20)	0.228 [0.69]	1.173** [0.03]
Years experience on-farm	0.017** [0.01]	-0.055*** [0.00]
Average years of education of rest of household	0.004 [0.71]	0.009 [0.34]
Average on-farm experience of rest of household	0.033*** [0.00]	0.018*** [0.01]
Number of people over 65 in the house	0.047 [0.92]	-0.399 [0.23]
Number of children (under 15)	0.525 [0.15]	0.110 [0.74]
Number of children (under 7)	-0.676 [0.23]	-0.251 [0.63]
Unearned income	0.001** [0.05]	0.001* [0.06]
Number of rooms in the family home	-0.381*** [0.00]	0.140 [0.12]
Spanish speaking	-0.553*** [0.00]	-0.065 [0.72]

Dependent Variable: participation	Random Effects Multinomial Logit	
	Unpaid farm work	Off-farm work
Home owners	-0.067 [0.81]	-0.290 [0.23]
Farm capital per adult	0.000 [0.86]	-0.001*** [0.00]
Farm size per adult	0.001 [0.98]	0.001 [0.87]
Change in farm size (1991-1994)	-0.016 [0.39]	-0.014 [0.46]
Inverted Simpson Index	-0.032 [0.73]	-0.181** [0.04]
Proportion of farm output sold on the market	-0.816*** [0.00]	-1.229*** [0.00]
Maize	0.383 [0.18]	0.273 [0.33]
Potato	0.185 [0.41]	0.440** [0.04]
Rice	-0.546** [0.04]	-0.705*** [0.01]
Banana	-1.055*** [0.01]	-0.273 [0.44]
Importance of local non-agricultural income sources	-0.186 [0.70]	1.141** [0.01]
Difficulty of access (distance)	0.015 [0.86]	-0.015 [0.85]
Difficulty of access (time)	-0.049* [0.10]	-0.039 [0.16]
Farm tenure (higher means less secure)	-0.080 [0.14]	0.032 [0.51]
Coastal region	0.107 [0.70]	-0.113 [0.66]
Rainforest region	0.715*** [0.01]	0.184 [0.47]
No. of men relative to women in area - men	-0.013 [0.63]	-0.010 [0.58]
No. of men relative to women in area - women	1.192*** [0.01]	1.180*** [0.01]
Migrated into area to get married	-1.147* [0.06]	0.493 [0.31]
Intercept	1.382 [0.16]	0.666 [0.47]
N	2453*	
Groups	1091	
Log Likelihood	-1759.9	
*this model includes some households not included in the other models: see Appendix 4.2		

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported. The excluded group is those who farm for returns, and do not work for wages off-farm.

Few coefficients change substantially in the random effects multinomial logit model of Table 4-6, compared to Table 4-5. The education variable is included non-linearly in this model, and it is clear that the effect is strongest for the very highly educated. Also, the farm size variable becomes insignificant, and the local ratio of men to women becomes significant positive for off-farm work for women. However, the difficulty of estimating the random effects model (it is extremely slow and reaching a solution is by no means guaranteed for large models), mean that it is not estimated for the polychotomous models of Table 4-7, and the assumption of non-independence of errors *within* households is the only concession made in Table 4-7 to allow for the intra-household simultaneity in decision making.

4.2.3.3 *Model c) Polychotomous dependent variables*

In Table 4-7 the number of values of the dependent variable is increased from three to six. This is because the category of 'off-farm work' subsumes that of both non-agricultural work and farm wage work, which are interesting to separate for policy reasons (Reardon *et al.* 2001).

Table 4-7: Multinomial logit individual participation (6 categories)

	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
Credit	0.013	0.207	-0.133	-0.741**	0.208
	[0.95]	[0.34]	[0.60]	[0.02]	[0.39]
Age	-0.088***	-0.002	0.103***	0.052	0.091**
	[0.00]	[0.94]	[0.00]	[0.20]	[0.03]
Age^2	0.001*	0.000	-0.001***	-0.001	-0.001**
	[0.09]	[0.46]	[0.00]	[0.24]	[0.05]
Attending school	1.928***	1.441*	0.896	0.479	1.476*
	[0.01]	[0.06]	[0.25]	[0.65]	[0.10]
Spouse	3.857***	-0.071	-0.403	0.846***	0.631**
	[0.00]	[0.74]	[0.14]	[0.00]	[0.01]
Other male	3.991***	0.752**	1.242***	-0.797*	-0.09
	[0.00]	[0.02]	[0.00]	[0.10]	[0.83]
Other female	4.677***	0.339	2.034***	1.525***	1.618***
	[0.00]	[0.48]	[0.00]	[0.00]	[0.00]
Years of education	-0.017	-0.036	0.088***	0.04	0.022
	[0.44]	[0.11]	[0.00]	[0.15]	[0.39]
Years experience on-farm	0.015**	-0.023***	-0.055***	-0.061***	-0.052***
	[0.02]	[0.00]	[0.00]	[0.00]	[0.00]
Average years of education of rest of household	0.047*	-0.028	0.069**	0.022	0.026
	[0.08]	[0.26]	[0.02]	[0.50]	[0.38]
Average on-farm experience of rest of household	0.021***	0.001	0.011	0.004	0.000
	[0.00]	[0.87]	[0.27]	[0.66]	[0.97]
Number of people over 65 in the house	0.257	0.181	-0.041	0.033	0.241
	[0.22]	[0.47]	[0.86]	[0.90]	[0.36]
Number of children (under 15)	0.006	0.023	0.012	-0.063	-0.066
	[0.93]	[0.76]	[0.88]	[0.48]	[0.41]
Number of children (under 7)	0.012	-0.001	0.027	-0.012	0.137
	[0.91]	[0.99]	[0.82]	[0.93]	[0.23]
Unearned income	0.001	0.001	0.001	0.001	-0.001***
	[0.11]	[0.14]	[0.16]	[0.15]	[0.00]
Number of rooms in the family home	-0.401***	-0.016	0.056	0.040	0.240**
	[0.00]	[0.87]	[0.62]	[0.73]	[0.01]
Spanish speaking	-0.350*	0.027	-0.092	-0.046	0.419
	[0.10]	[0.90]	[0.69]	[0.86]	[0.10]
Home owners	-0.046	-0.032	-0.354	-0.410	-0.414
	[0.87]	[0.91]	[0.22]	[0.19]	[0.11]
Farm capital per adult	0.001	0.001	-0.001***	0.001	-0.001***
	[0.76]	[0.18]	[0.00]	[0.31]	[0.00]
Farm size per adult	-0.026	-0.047	-0.052**	-0.068**	-0.039
	[0.26]	[0.23]	[0.04]	[0.03]	[0.15]
Change in farm size (1991-1994)	0.052	-0.014	-0.018	0.027	-0.056
	[0.23]	[0.80]	[0.75]	[0.66]	[0.36]
Inverted Simpson Index	-0.137*	-0.085	-0.373***	-0.162*	-0.364***
	[0.08]	[0.32]	[0.00]	[0.09]	[0.00]
Proportion of farm output sold on the market	-0.886***	-1.473***	-1.648***	-0.879**	-1.321***
	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]
Rice	-0.260	-0.742***	-1.084***	-0.626**	-0.533
	[0.31]	[0.01]	[0.00]	[0.05]	[0.13]
Maize	0.249	0.449*	0.123	0.033	0.900***
	[0.25]	[0.06]	[0.62]	[0.91]	[0.00]
Banana	-0.172	-0.345	-0.218	-0.535	0.23
	[0.62]	[0.46]	[0.60]	[0.26]	[0.55]
Potato	0.280	0.102	0.026	-0.216	0.942***
	[0.23]	[0.71]	[0.92]	[0.50]	[0.00]
Importance of local non-agricultural income sources	0.388	1.081**	0.397	0.596	1.861***
	[0.37]	[0.02]	[0.52]	[0.31]	[0.00]
Difficulty of access (distance)	-0.059	-0.149	0.042	-0.149	0.178*
	[0.50]	[0.13]	[0.68]	[0.16]	[0.07]
Difficulty of access (time)	-0.049	-0.083**	-0.003	-0.082**	0.017
	[0.10]	[0.01]	[0.92]	[0.03]	[0.57]
Farm tenure (higher means less secure)	-0.089*	0.008	0.059	-0.026	-0.076
	[0.08]	[0.89]	[0.32]	[0.70]	[0.25]
Coastal region	0.273	0.677**	-0.273	-0.227	-0.494

	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
	[0.33]	[0.03]	[0.37]	[0.53]	[0.19]
Rainforest region	0.431*	0.475	-0.367	0.339	0.271
	[0.10]	[0.10]	[0.25]	[0.31]	[0.45]
Unpaid labour exchange is customary	0.472**	-0.159	0.427*	-0.203	0.492*
	[0.02]	[0.50]	[0.07]	[0.43]	[0.06]
Intercept	0.090	1.389	-1.497	0.073	-3.910***
	[0.93]	[0.20]	[0.18]	[0.95]	[0.00]
N	2344	2344	2344	2344	2344

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported. The excluded group is those who farm for returns, and do not work for wages off-farm.

Furthermore, it is also interesting to divide the self-employment category into sole traders and employers, since that leaves us with four off-farm work categories which conform with ILO international labour force categories. Thus the six values of the dependent variable are: on-farm paid, on farm unpaid, farm wage earner, plus the three non-agricultural categories: non-agricultural wage earner, self-employed sole trader and self-employed employer.[18]

These results add substantially to what we have already seen. The four off-farm categories all exhibit the familiar concave age profile. Non-agricultural wage labour is a category to which (female) spouses do not appear to have access: female spouses in the off-farm sector tend to enter self-employment, with a slight leaning towards being sole traders rather than employers. Male children in the off-farm sector tend to enter wage labour, while female children off-farm tend to enter self-employment as employers or to enter non-agricultural wage labour.

Having more education appears to push people into the non-agricultural wage category and away from unpaid on-farm work. If others in the house are more educated, then unpaid farm work is more likely, as is non-agricultural wage labour. Having on-farm experience increases the chances of doing unpaid farm work and reduces the chances of doing any off-farm work in any sector. If others have more on-farm experience one is more likely to do unpaid farm work.

Unexpected unearned income reduces the (relative) likelihood of entering the larger self-employed category, though it is negative and close to significant for all off-farm categories.

People in large houses tend not to enter the unpaid farm worker category, but do tend to enter the self-employed employer category. Given imperfect capital markets,

[18] These regressions were also explored with a seventh category with 697 new people – those doing only work within the home. Results for the six above categories were not affected. For the seventh category - home production - the main factors leading to entry into this category were: being at school, being female, being inexperienced on the farm, being in a house where others are relatively more experienced on farm, having less unearned income and living in more remote areas.

increased wealth (proxied by the number of rooms) should lead to increased larger scale self-employment, but there is some possibility of endogeneity as far is this variable is concerned.

The farm capital variable, on the other hand, is significantly negative for both the non-agriculture wage category and the self-employed employer category. Farm size is negative for categories four and five – non-agricultural wage and sole trader self-employed. Thus, being from a small farm appears to reduce the size of the 'self-employed' business.

Being from the least commercial farms also leads to entry to both the non-agricultural wage and employer categories, suggesting simultaneous planning of off and on farm strategies.

The diversification variable is generally significant in these estimations, except for the farm wage category. This latter result suggests that there probably is an element of portfolio diversification in off-farm work. Over time, returns in on-farm work and off-farm wage work on farms are probably more highly correlated than returns on on-farm work and those in non-farm work. Since on-farm diversification is associated with reduced non-farm work, but not reduced off-farm wage work, it does seem to be legitimate to speak of non-farm work as part of a household diversification strategy, and not simply an extra source of income for when returns are higher than farm returns.

The coastal variable is significant for the first time. The coefficient is positive for entry into paid farm labour.

Being from the most remote areas appears not to be conducive to people entering agricultural wage labour and/or being a sole trader, whereas being from a non-agricultural area increases entry to agricultural wage labour and larger scale self-employment. The minka (customary unpaid labour exchange) variable is positive for unpaid farm labour (unsurprising) and for the non-agricultural wages category (surprising). In an effort to explore this result the sample was split into Spanish speaking and non-Spanish speaking households. Minka has a positive effect on entry into unpaid labour for non-Spanish speaking households but it also has a positive effect on entry into non-farming wage labour for Spanish speaking households, so the reported effect in Table 4-7 conflates these two separate effects.

4.2.3.4 *Model c) Two extensions*

The operator/spouse relationship has been modelled in the bivariate probit models and reported in Table 4-2. In the logit and multinomial logit models of Tables 4.3-4.7 the operator/spouse and operator/household framework has been replaced by modelling participation decisions at the individual level and accounting for within-household interactions and household individual effects, firstly, by including on the right hand side relevant variables at the household level (e.g. average education of others in the house, or average on-farm experience of others in the house) as well as, secondly, by modelling the error process either as being independent between households but dependent within them, or by using random effects models at the household level.

In these models, though, we have seen that one restriction imposed is that coefficients (except for intercept dummies) are assumed to be the same across household members. This restriction is not imposed in the Bivariate Probits of Table 4-2, but is also implicit in the household level model of Table 4-1. In the final models of this section, this restriction is relaxed. Two types of models are run: multivariate probit models, and multinomial logits with 'slope dummies' (three dummies – for operator, spouse and adult child – each multiplied with all the other included independent variables). Both of these models, therefore, allow flexibility in the size and signs of coefficients for three types of family members – operator, spouse and other (nearly always a child). The extra information gained from this relaxation is interesting but the multinomial tables, at least, are very long, so all the tables have been confined to appendices (in Tables 4-A-1 and 4-A-2).

Summarizing the key results – it is possible, firstly, to claim that both of these extensions of the 'individual within the household' model show that intra-familial interactions are indeed important. In the multivariate probits, the (econometric) non-separability of operators' and spouses' decisions, already noted in the bivariate probit models of Table 4-2, is confirmed. In the present models (Table 4-A-1) there is also correlation found between (female) spouses' and children's decisions, and none between those of the operator and the child. There is no noticeable difference in the strength of these correlations according to the sex of the child. This pattern of intra-familial correlation is sufficient, however, to enable one to claim that the bivariate probit models of Table 4-2 (which are quite typical in the literature, at least for developed countries) are mis-specified, since the spousal participation decision cannot, it appears, be modelled separately from EITHER the operator's or her children's decision(s).

From the flexible (and large) multinomial logit models of Table 4-A-2, examination of results casts light on a number of features that have so far been under-examined. For instance, we can elaborate the influence of the cost of diversification variable. We note from this Table that the use of off-farm work as a response to not being able to diversify on-farm is something more likely to be done by the operator (in the form of self-employment or non-farm wage work) and adult child (also in the form of self-employment or non-farm wage work), rather than spouse. For the adult child, unlike in Table 4.7, diversification does affect entry to farm wage labour; but for the operator it still does not, suggesting that diversification strategies are mainly managed by the operator.

On farms that do not sell crops on the market for cash, both operator and spouse are likely to work off-farm in all categories, wage work plus self-employed. Their adult children, also, are more likely to do farm wage work. Household cash flow, therefore, appears to be mainly the concern of all family members.

Similarly, this estimation contains more information on the 'number of rooms' variable. It tends to be positive for off-farm participation in a variety of sectors for operators and adult children, and – for the first time in the regressions – negative for spouses. These results are possibly more consistent with an exogenous interpretation for this variable than for an endogenous one (how would a spouse staying at home increase house size?), but some endogeneity is likely (otherwise – if the variable only captured the effects of capital market imperfections, as intended – it should be significant for self-employment, only, not wage work).

Education comes through strongly for the operator (for non-farm wage work, which is less common than farm wage work) and for adult children (keeping them out of agricultural wage work). Education improves spouse's chances of getting into self-employment.

Unexpected unearned income is significant for the spouse in most categories, and for operators only in category six (and, with a positive sign, category 1!), and not at all for adult children. This suggests that the spouse's shadow wage is probably higher – all else equal – than that of the operator and adult children.

The family 'bargaining' variable is significant for the spouse: if she marries in, she tends to work for wages in a non-agricultural occupation.

For the adult children, the effect of being in a non-agricultural area is strong. There are not many non-agricultural areas (a little less than 10% of segmentos), so it is hard to gauge this effect, but from the evidence it seems likely that adult children's off-farm chances of employment are very much location driven than those of their parents. The signals from the location variables are weaker and more mixed for operators and spouses. This variable was also strong for both operator and spouse in Table 4.2, but in the larger model that effect has disappeared.

This is the first estimation where having children has shown through as significant. Having children reduces spouses' entry to sole trader self-employment, their most common category. The lack of effect of young children on choices has been one notable feature of these results.

It is also the first estimation where the change in farm size (from 1991 to 1994) has come through – for adult children. The sign is significant positive for entry into unpaid farm work, but also non-agricultural wage work and large scale self-employment. This is the first hint of any complementarity between off-farm work and on-farm assets. It suggests that for households on some expanding farms (there are not very many), income growth could occur both through farm growth and the non-agricultural sector.

Finally, if this model is re-run using actual values for the averaged quantitative household variables (number of rooms, unearned income, farm size and capital, and change in farm size) instead of per adult averages, there are few changes and full results are not reported. The most notable change is that total farm size becomes positive significant for the spouse to enter category four (paid wage work outside agriculture.

4.3 Discussion of main points of participation results

The increase in information that the later estimations provide comes mainly from a) a decomposition of job options (from two to six values for the dependent variable) and b) an increase in the number of people modelled within the family (from one 'average' person, to operator and spouse, to all adults over 15 years of age) – the choice of estimator (random effects or not), while relevant statistically, rarely changes the later substantive results (except for household tenure variable). Depending primarily on the amount of options in the dependent variable, and on the number of people modelled within the family, results change and certain subtleties are gained or lost. Some examples of the gains in information from the larger models are:

- Other things equal, relative to operators, adult male children tend to work off-farm for wages; younger females tend to work for both wages and self-employment. Spouses tend to work off-farm in self-employment.
- A higher average level of education tends to increase the likelihood that someone in the household will work off-farm. When the job options are disaggregated, higher education tends to reduce entry to agricultural work and raise entry to non-agricultural wage labour. The main effect on the operator is to propel him into wage labour or large-scale self-employment (this effect does not show through in the relatively aggregated bivariate probit models). The main effect on the spouse is greater entry into small-scale self-employment. The main effect on the adult children is to reduce the chances of entry to agricultural wage labour.
- One motive of off-farm work appears to be to manage risk. This can be seen from the results in Table 4.7 (on-farm diversification reduces entry to all types of off-farm work, except for agricultural wage work, the returns to which are likely to be correlated with on-farm family work) and Table 4.A.2 (where basically the same result holds, but only for operators: there is no connection whatsoever between on-farm diversification and spousal decisions, while for adult children in the household the negative relationship holds between on-farm diversification and all categories of off-farm work, including agricultural wage work). The suggestion is that part of the reason, at least, for off-farm work is to manage diversification, and that this reason applies mainly to the operator, otherwise the effects would be similar across all types of off-farm work. If this really is the case then it is likely that having earnings from off-farm work will improve allocative efficiency, as hypothesised in the Introduction to this thesis. This hypothesis is tested, at the household level, in Chapter 6.
- If either spouse or operator has paid off-farm work, then the household is less likely to produce cash crops on the farm. This relationship also holds where adult

children are concerned, suggesting that household cash-flow is mainly managed by all family members.[19]

- Having young children tends to reduce entry into category five (being a sole trader) for the spouse, but has no other effect on participation.
- The effect of extra unearned income tends to be felt almost wholly by the spouse – reducing her entry into both unpaid farming and wage labour. The effect on adult children's decisions is non-existent, while there is a moderate effect on operators' decisions. All else equal, the spouse appears to have a higher shadow value of home time (or greater control of unearned income).
- Wealth (as measured by the size of the home) tends to be associated with a higher likelihood of operators' being in self-employment, and a lower likelihood of spouses being in wage work.
- The farm capital and farm size variables are consistently significant negative in all the estimations, particularly for the operator, clearly suggestive of negative interdependencies between investment on the farm and investment off it, and of the operator's role in these decisions.
- Interdependencies in decision making regarding participation appear to be strong between operator and spouse, and spouse and children, but not between operator and adult children. In particular, the spouse's decisions seem to be related to those of all other family members.
- Community level demand variables tend to have stronger effects on adult children than on the operator or the spouse, though the evidence for this is slightly thin.
- Adult children on growing farms are more likely to work off-farm rather than on it. This is a little surprising, and is suggestive – perhaps – of a long-term strategy in some households of long range planning of household income diversification, or movement from the land for some children.

[19] The 'receipt of credit' dummy variable was inserted in exploratory specifications for the models in Table 4-7 and A2 (the polychotomous multilogits). It is only significant once – for entry to category 5, with – perhaps surprisingly - a negative sign in Table 4-7. The A2 specification result suggests this negative sign is not person specific but appears to be spread (insignificantly) over all family members: thus entry to category 5 – sole trader self-employment – may be a recourse to any family member when credit that might be otherwise useful is unavailable.

- Being from a larger family tends to increase the probability of someone in the household working off-farm.
- If an operator is more experienced on the farm, a spouse is more likely to work off it. The opposite is not true. Also, spouse's participation is dependent on operator's age; again, the opposite is not true.
- Operators on rice farms do not tend to work off-farm. Operators on maize farms often work off-farm for other farmers. Spouses on banana farms may work off-farm, and adult children on banana and potato farms may work off-farm.
- Women who marry into an area tend to work off the farm for non-agricultural wages. Nobody who marries in is likely to do unpaid farm work. In areas with more men than women, men are generally likely not to work off the farm and women are more likely to do unpaid work on the farm, and to work off farm. Neither of these variables appear as if they can be interpreted as bargaining variables, even though the second – certainly – is used as one in some Collective model tests (Chiappori *et al.* 2001).
- Non-Spanish speakers tend to work on the farm un-paid, suggesting possible differing intra-familial allocation patterns for differing ethnic groups (in accord with the evidence of the last chapter). Non-Spanish speakers generally are less likely to work off-farm: but once unpaid on-farm work is controlled for, then there is no difference in participation according to language group.
- Being in an area with voluntary labour exchange increases the chances (for non-Spanish speakers) of doing unpaid work on one's own farm, and increases the chances (for Spanish speakers) of doing non-agricultural wage labour.

The importance or otherwise of these conclusions is context dependent, and a measure of their theoretical and economic significance can only be made within a structural model. However, the objective of this chapter has been to explore the shallower waters of the simpler reduced form participation models and test their sensitivity to certain aspects of model specification. It would seem that these aspects are fairly important.

4.4 Overall discussion

Participation and hours models (reported in the appendix to Chapter 4) were estimated for three different decisions making units – the household, the farm couple and the individual adults within the household. For the latter, three different sets of participation models were estimated – with dichotomous dependent variables (on– and off-farm work), with trichotomous dependent variables (on-farm paid, on-farm unpaid and off-farm work) and with polychotomous dependent variables (on-farm paid, on-farm unpaid, off-farm agricultural wage, non-agricultural wage and non-agricultural self-employment, either with or without paid employees), as well as a couple of brief extensions new to the literature (multivariate probits and flexible multinomial logits with differing coefficients for family members). Different error processes were modelled, and an instrumental variable version of the household model was also estimated. A particular selection mechanism was also specified for the multinomial hours estimations in the appendix.

The objective of this step-by-step procedure has been to try to gauge the importance of specification for results from reduced form participation and hours models. It is found that the simplest specification (the household participation model) yields a set of results that are rarely substantially contradicted by any of the later models, but often need some amendment and 'fine tuning'.

In the simplest specifications there are significant variables at community, farm and household level, and there appear to be interesting interactions between on-farm investment strategy and off–farm labour strategies (the farm capital coefficient is invariably negative), and between cash-flow and on-farm risk management strategies (the diversification and proportion of crops sold variables also tend to be negative). The participation instrumental variable estimations suggested that these three variables may be endogenous with participation (but the quality of the instruments used is weak, and exogeneity is likely on other grounds). Other interesting results of the simple household model relate to gender (significant effects), on-farm experience (strongly significant) and farm and home tenure (usually insignificant).

The more individualised models generally confirm or complement these results. They reveal a substantial amount i) about intra-family sectoral allocation (e.g. spouses move into self-employment, younger males into wage labour, younger females into the non-agricultural wage sector and self-employment), and ii) about the mechanics of intra-family decision making (e.g. on-farm / off-farm risk is mainly managed by the operator, while on-farm / off-farm cash flow decisions are managed by all members: more generally, on-off-farm allocation is decided by operator and spouse simultaneously, or spouse and children together, but not simultaneously by operator and children).

Some other notable conclusions from the individualised models are that unexpected off-farm income affects mainly spouses' decisions and not the adult children's, while own education affects the operators', spouses' and children's decisions, and tends to reduce the chances of adult children working for agricultural wages. The 'push' from more education generally is towards non-agricultural wage work but for the spouse it is towards self-employment.

Women who marry into an area tend to work mainly for wages but the effects of the local male/female ratio are erratic, and this variable is probably uninterpretable in the context. Spouses' participation is affected by operators' age and experience but operators' participation is not affected by the education or experience of the spouse.

The hours regressions in the appendix add little, but two points are worth mentioning. Having children appears to increase hours worked in certain categories (even though participation is not generally affected by having children). Also, there are some farmers working long hours in large scale self-employment whose farms are getting bigger over time. This suggests some use of off-farm income (for a very small number of people) in expanding the farm. Generally on and off-farm work are substitutes, but for a few people they appear to be complements.

The community level 'demand' variables, particularly the importance of agriculture in the community and the relatively easy availability of public transport, tend to work in the expected way in the more aggregated models. These variables seem especially important for the adult children.

The wealth indicator (number of rooms) is somewhat problematic, likely to be both endogenous (a result of past efforts) and, in some cases, exogenous (serving to overcome capital market imperfections and increase the likelihood of being in larger scale self-employment, as well as indicating higher unearned income, reducing hours and, possibly, participation).

How the errors are specified and the ensuing choice of estimator are statistically relevant (in hardly any case, for example, was the unimportance of individual household or segmento effects not rejected), but these choices are not important, at least for this dataset, in terms of changing coefficients dramatically, or changing significances. The primary methodological conclusion of the section is that univariate or bivariate dichotomous models of household participation, which are by far the most common in the literature, provide a reasonable picture of the main determinants of household off-farm participation, but that more disaggregated modelling of participation and hours can reveal additional information that may be of considerable interest, both academically and for policy makers.

5

Separability, the household and off-farm labour

Introduction

In this chapter, aspects of separability in labour allocation decisions are examined. Separability, in this context, means that home consumption and farm production decisions are separately made. Production decisions are determined by output prices and input costs. Consumption decisions are determined mainly by the prices of consumption goods, earnings and 'unearned' incomes, household characteristics (e.g. the presence of young children) and tastes. The chapter focuses primarily on two possible breakdowns in this kind of separability. First, if individuals are constrained in their off-farm decisions, such constraints are likely to lead to modification of behaviour on the farm. Second, if individuals are constrained in their ability to hire in farm labour because of monitoring costs, there are direct implications for on-farm work patterns as well as indirect implications for the most efficient distribution of farm sizes.

From the literature, the main implication of either kind of non-separability in the farm household labour allocation process is that a dependence can arise between consumption and production decisions. If consumption variables – such as wealth, household size and composition, non-earned income and the prices of consumption goods – are known to affect on-farm production decisions for those households already working off the farm (or those households already hiring in), separability can be said to have broken down.[20] The implications and formal consequences of this will be discussed in section 5.2, and right throughout this chapter. In some cases (when labour markets are rationed in hours for those working off the farm), at least one other imperfection in land, capital or insurance markets, as well as the imperfection in the labour market, may be needed for such effects to arise.[21] The question of what effects

[20] Benjamin, 1992, distinguishes weak and strong separability. In weak separability hiring in and hiring out wages may differ, but consumption variables do not affect production.

[21] Kevane, 1996, works through the fairly knotty comparative statics for the labour land ratio for 6 cases of 'double' market imperfections – labour and land rental, labour and credit, labour and risk, land rent and credit, risk and land rent, risk and credit. 3 imperfections at once, he suggests, is too difficult to handle. 1 market imperfection

on on-farm production might be expected from imperfections in capital and insurance markets (for those individuals/households hiring out or in) is discussed in section 5.1.2.

The households (and individuals) focused on for the rest of the chapter are those farm households that are dividing their time between on and off-farm work whether they hire in or not, or else those households that work on the farm only and hire in. It is assumed throughout the chapter that there are land market rigidities (so that off-farm labour rationing, for example, does not simply lead to an increase in farm sizes). This assumption is certainly acceptable for the Peru of 1994, where land sales and rental were extremely rare.

These choices and assumptions are made (and refined below in some cases) because the tests used below to check for separability are conditional on the household (or individual) regime. If a household is in pure autarchy (neither hiring out nor in), for example, then consumption and production decisions are perforce simultaneous as per the standard version of the farm household model in Chapter 1, and the possible existence of barriers that prevent escape from pure autarky cannot be directly tested for. It is important to note, also, that testing for separability when the household are already participating off-farm is only one of many routes into an investigation of the constraints operating on farm households/individuals.

The main methodological advances in this chapter are that, as with the last chapter, the issue of testing at the household level or the individual level is addressed (in section 5.5); also, a new test is proposed (in section 5.3) to distinguish rationing from non-linear transaction costs as causative factors in the breakdown of separability.

The chapter begins with some examples of non-separabilties, a short literature review, as well as some discussion of difficulties with reduced form tests. A model is then introduced, and reduced form tests carried out for separability in working off-farm and in hiring in. In the last section, an econometric model for dealing with household heterogeneity in separability is introduced, then estimated and finally results are discussed.

---

can be 'got round' if it is the only one (e.g if off-farm labour is rationed, then farm size could be expanded). In Peru, land markets were extremely sluggish at least until the 1995 reforms so land allocations are assumed fixed in this chapter.

5.1 Separability and non-separability in labour allocation decisions

The issue of separability has been an important issue in farm household modelling since the earliest models (Singh *et al.*, 1986; Taylor and Adelman, 2002). In a separable regime, farm production is determined by prices, including the price of labour, as well as productive resources and technology. The farm household model, as discussed in Chapter 1, refers to a separable regime.

However, in many middle income and developing countries separability may not hold. These cases have also been discussed in Singh *et al.*, (1986); Taylor and Adelman (2002), as well as in de Janvry and Sadoulet (2004). Section 5.2, below, formally extends the household model to allow for non-separabilties.

De Janvry and Sadoulet (2004) give an example of the effects of non-separabilities on farm household decision making. Joint failures in the markets for food and for labour off the farm may affect responses to changes in prices of cash crops. A rise in the price of cash crops, for example, cannot lead to an increase in supply if farmers are constrained to grow enough food for their families' survival and are constrained from hiring in labour due to weak labour markets or very high transactions costs. In this case only technology improvements or better inputs can raise production of cash crops. Lofgren and Robinson (1999), in another example, show that large transaction costs and the resulting regime switches strongly dampen responses to increases in cash crop prices on the international markets. Conversely, if the markets for imported consumption goods fail, then supply response of cash crops is also weakened (there is nothing to spend the money on!). In another related example, technological change in food production, for example, can raise supply elasticities of cash crops (Dutilly-Diane *et al.*, 2003). Cheaper imports following a tariff reduction incentivizes production of cash crops (de Janvry and Sadoulet, 2004). Taylor and Adelman (2002) also give an example (referred to in Chapter 1) where a policy induced increase in the price of rice led to *less* rice appearing on the market because the increased income from rice enabled households to consume more of their own rice and less inferior staples.

Separability in the labour market (for hired-in labour) is also a key issue in the ongoing debate over intensity of factor usage and the putative inverse relation between yield and farm size. It is argued (at least since Berry and Cline, 1979) that high monitoring costs of hired-in labour make family labour more efficient than hired labour, leading to the inverse yield-size relationship for farms above the size of a family farm. Thus, the most efficient size distribution is the Chayanovian one (Chayanov, 1926) compatible with a large number of family farms needing little hired labour. Barrett (1996) also makes the case for the efficiency of this kind of distribution, but based on price risk in agricultural output markets. The inverse farm size productivity relation has been found in many studies (Berry and Cline, 1979; Benjamin, 1995; Barrett, 1996, Lamb, 2003), but evidence is difficult to evaluate because of land quality issues (Benjamin, 1995; Bhalla and Roy, 1988). Kevane (1996) finds the opposite relationship, which he explains as being due to a positive relationship between wealth and yields consequent on failures in capital and insurance markets.

An issue which is rarely discussed in the literature is the choice of whether to test for separability at the individual or at the household level. If there is off-farm labour rationing, for example, it could affect entire communities, whole households[22] and / or certain individuals within households. The effects on individuals could either be externally or internally induced (e.g. market discrimination against women, or intra-familial decisions about whether women can or should work off the farm). Lambert and Magnac (1994), for example, find that 90% of men in their sample in the Ivory Coast are constrained, while only 50% of women are. Morrugara (1998) finds that separability may hold less for women than men, if women tend to do most of the on-farm monitoring of hired-in labour. However, reduced form tests have not generally been estimated at the level of the individual, but rather at the level of the household, with compositional variables included (e.g. average gender). It is possible that there are gains from estimating at least some of these regressions at the level of the individual. This issue will be discussed in section 5.5.

5.1.1 Testing for non-separability

Although there is a long history of testing for separability of labour supply and consumption in developed countries (Abbott and Ashenfelter, 1976; Barnett, 1979; Blundell and Walker, 1982; Browning and Meghir, 1991 to name but a few), and for testing for separabilities between different types of consumption goods (Deaton, 1988;

---

[22] ......because of location say, or ethnicity - Vakis, *et al.* (2004) find that non-Spanish speaking households are more likely to be rationed in Peru in 1997 than Spanish speakers.

Deaton *et al.*, 1989) it was not until Benjamin's 1992 article that testing for separability in labour allocation began to become fairly common for farm households (previous tests had been carried out by Pitt and Rosenzweig (1986) – using a very specific test involving illnesses in the farm family and farm profits – and Arayama, in unpublished work in 1986, whose tests are more similar to Benjamin's). Most reduced form tests of separability involve regressing the total amount of farm labour or family farm labour household consumption variables (e.g. number of adults in the household, or unearned income), controlling for variables affecting farm output. If the consumption variables are significant, then separability is deemed to have broken down. This result follows immediately from the definition of separability of labour allocation decisions in the farm household context – consumption variables not affecting production decisions.

Benjamin (1992) and Bowlus and Sinclair (2003), using this 'reduced form' version of the test, find no breakdown for Java and China; while Arayama (1986) and Grimard (2000) find breakdowns for the Japan and the Ivory Coast. Jacoby (1994) and Skoufias (1994), using a slightly more structural approach, estimate marginal production in agriculture in Peru and India, respectively, and test for its equality with the off-farm wage. In both cases they reject separability. Lambert and Magnac (1994) and Bhattacharyya and Kumbhakar (1997) do something similar for the Ivory Coast and Bengal, but allow for heterogeneity among households by estimating the household specific standard error for each marginal product and then classifying males and females within households according to likely regime.

Other approaches also allow for heterogeneity among households, while at the same time assuming that market position is a good indicator of separability status (e.g. if a household has access to credit it is not credit constrained; if a household is working off-farm it is not constrained in working off-farm). These include Feder, Lau and Lin (1990), Carter and Olinto (2000), Carter and Yao (2002), Sadoulet *et al.* (1998) and Dutilly-Diane *et al.* (2000). Vakis *et al.* (2004) also allow for heterogeneity but in this case, unlike the above references, the market position of the household (participating or not) is not a good indicator of lack of separability and an endogenous switching model is used to identify household regimes (separable or non-separable). In their case, some households working off-farm are found to be constrained in the amount of off-farm work they can do. Finally, Henning and Henningsen (2005) allow for heterogeneity in wages within households, which leads to a lack of separability on the assumption that those household members who will be better paid will work off-farm first so that hourly wages are decreasing in household off-farm time worked – that is, wages are non-linear at the household level. As noted in Chapter 1, section 1.1.2, non-linearities in off-farm returns always lead to the breakdown of separability.

If separability is rejected, the sources of any non-separabilities found are rarely examined. The assumption is generally made that they are due to off-farm labour rationing (Hart's 1986 study of a Javanese rice village has been influential in this regard), with a tacit assumption generally of weakness in land markets or other related imperfection, or else that they are due to transaction costs.

5.1.2 Market imperfections and on-farm production

Before going on to test for non-separabilities in the Peruvian labour market, this sub-section very briefly addresses potential confounders to the usual interpretation of the rejection of the null for consumption variables in the reduced form tests referred to above. These are imperfections in the capital and risk markets. There is also a brief teasing out of some of the reasons why non-separabilities might appear in the labour market.

Regarding imperfections in the capital market (Carter and Olinto, 2000), if high fixed costs of lending, for example, reduce the likelihood of asset poor farmers receiving loans – the usual reason for capital market imperfections – then effects may arise in on farm decision making (e.g. crop mix and yields on farm) and labour time on and off the farm. The main problem with regard to testing for labour separability is that both earned and unearned income are likely to be used to get round the capital market imperfections (this is true also for indicators of wealth). For example, Reardon *et al.* (1994) and Kevane (1996) find in that off-farm income and wealth, respectively, influence on-farm productivity. Such productivity increases will increase returns to farming, probably lead to increases in on-farm working time (for those already working off-farm). However, even if the capital market is imperfect, for those households already participating off farm consumption variables such as family size should not affect on-farm allocations if the labour market works smoothly as long as farm capital is controlled for in the regressions, which is the case in the regressions below.

With regard to the management of risk, if formal insurance markets against income fluctuations are imperfect (as they usually are), then households are likely to insure themselves informally in a variety of ways.[23] In particular, more risky off-farm work (e.g. seasonal, secondary or non-insured) means that family characteristics may also affect the time spent working on farm (increasing it), as well as the specific tasks carried out on the farm (broadening the crop mix, or moving to less risky crops). In both of these cases, at least, non-separabilities occur, and household characteristics may be expected to influence total hours on-farm despite the fact that the household also has off-farm wages coming in. For separability testing of labour allocation decisions, it is important to include a farm activity diversification variable to (partially) control for risk management. Leaving it out could be important if crop diversification is correlated with household size or composition variables. Including this variable in the regression does not completely control for on-farm risk-management behaviour (the precise choice of outputs is not controlled for). Nevertheless, the inclusion of the risk diversification variable in the regressions below, as well as the 'main crop' variables, should be expected to control somewhat for on farm crop mix.

Returning to labour market non-separabilities themselves, the main focus in the literature has been on rationing and on transaction costs, including monitoring costs of hired labour, as explanations for non-separability. Lack of separability may be due to off-farm rationing resulting from seasonality of certain kinds of work (or rationing for other, perhaps sociologically based, reasons, preventing wages in certain occupations falling to clear markets). There may also be cultural reasons that constrain the labour market choices of households or individuals (e.g. due to the way family needs are culturally articulated some females or non-main persons within the household may not be able to work full time off-farm, for example, or operators may not work off-farm because their status is bound up with their being on the farm). There may also be

[23] One way - if decisions are Pareto efficient in the face of risk - is by mutual household insurance against shocks to individual earnings. In a Collective model, the Pareto weight would not be expected to depend on the realizations of individual income, say, for any given distribution from which such realizations might emerge (though it should depend on the parameters of the distribution itself). Thus, for instance, actual personal non-labour income shocks should be fully compensated for WITHIN the household (the same is true for the Unitary model) and so, conditional on total household expenditure, no individual's consumption should be affected by individual earnings shocks. The evidence for this is decidedly mixed (Udry, 1996, Dercon and Krihnan, 1996, Goldstein, 2001; and – at the village level - Townsend, 1994, Udry and Conley, 2004). In fact, on the available evidence (from Ghana), one's friends and one's extended family (Goldstein, 2001) appear as if they may be more important providers of insurance than are one's own immediate family. This suggests that a wider and more complex framework is needed for an analysis of risk management than a household model.

dynamic factors pertinent to either individuals or households, which lead to apparent lack of separability in a static model (Lambert and Magnac, 1994).

If rationing does not apply, then it is assumed that transaction costs must exist for separability to break down, but these transaction costs themselves are rarely modelled. In general, transaction costs may be fixed, proportional or non-proportional. In the first case they affect the agents' position on the market (whether they participate or not), but for those already participating separability is not affected. In the third case, of non-proportional transaction costs, entry or exit may also be affected (de Janvry and Sadoulet, 2004) but given participation (as a hirer or worker, buyer or seller) non-separability also occurs with this class of transaction costs not because they are transaction costs per se, but because of the non-linearities involved.

All of this means that reduced form tests that reject separability on the basis of consumption variables affecting on-farm hours do not always distinguish whether the alternative hypothesis (non-separability) may be accepted due to capital, risk or labour market imperfections. This is true also, at least as far as insurance markets are concerned, for the slightly more structural tests of Jacoby and Skoufias and Lambert and Magnac and Bhattacharyya and Kumbhakar, who estimate marginal products of agricultural labour and test for their equality with wages. Furthermore, there are severe data problems in estimating farm production functions, so severe that they render their results dubious (Schultz, 1990). It is these data problems that lead Vakis *et al.* 2004, to recommend that only reduced form models be used in testing for separability. But it is important to note that the trade-off involved in this choice is that, while using reduced forms of the farm household model makes it easier to detect whether separabilities exist or not, it tends to make it less easy to categorise and to quantify the effects of specific non-separabilities.

5.2 Separability in the labour market – the model

The farm household with k partners is assumed to solve the following variation of the model in Chapter 1:

$$\max \sum_{i=1}^{k} \mu_i U_i \qquad (5-1)$$

subject to

$$G(x,r) = Y \qquad (5-2)$$

$$T_L - X_L - X_L^s - C_L + X_L^h \geq 0 \qquad (5-3)$$

$$T_L = \sum_{i=1}^{k} T_{Li}; X_L = \sum_{i=1}^{k} X_{Li}; X_L^s = \sum_{i=1}^{k} X_{Li}^s; C_L = \sum_{i=1}^{k} C_{Li} \qquad (5-4)$$

$$P_m C_m \leq P_c X_c + P_a(X_a - C_a) - (w_h X_L^h - g(X_L^h, z_L^h)) + \sum_{i=1}^{k}(w_i X_{Li}^S - f(X_{Li}^S, z_{Li}^s)) + E \qquad (5-5)$$

$$X_{Li}^s \leq \overline{X}_{Li}^s; X_{Li}^h \leq \overline{X}_{Li}^h \qquad (5-6)$$

The utility function is an expanded version of the caring Collective Model, where the sum of weights is 1, although this adds nothing essential to the Unitary model in this context. Production is a multi-output, multi-input production function (2), where sets of outputs are goods for market (C_m) and goods for home consumption (C_a), and leisure is C_L. The factors of production are labour (family and hired in) and the fixed factors r – capital and land. Production is assumed to be separable in variable inputs. The total household time constraint is embodied in (5 – 3), where X_L is total time in family farm labour, X_L^s is family time spent in off-farm labour, X_L^h is the total labour hired in. In equation (5 – 5), consumption (on the left) is limited by money from sales of market only goods, money from sales of goods that are consumed at home, the costs of hiring in (represented by the wages paid as well as the possibly non-linear function g(..), where z includes variables affecting transaction costs), the returns to working for wages – w – (reduced possibly by transaction costs, represented in the possibly non-linear function f(…)) and unearned income from other sources (remittances etc.) designated by E. Possible labour rationing of off-farm or hired-in individual labour is given by equation (5 – 6). In solving the on-farm labour supply equations based on the first order conditions of this problem it can be shown (Lee, 1998) that non-linearities in f or g are sufficient for consumption variables to affect on-farm labour (family labour in the first case; total labour in the second). Likewise, it can be shown (Vakis *et al.* 2004) that rationing of off-farm labour at the going rate, if it exists, leads to the same result.

Farm production and off-farm wage work are arranged within the household to maximize household returns, given the individual utility functions and the distribution of bargaining weights within the family. The assumption made by Henning and Henningsen (2005) that those with the highest wages work off-farm first is not necessarily appropriate in this setting. Assuming indifference between working on or off the farm, the person who provides the largest household net gain off-farm will be the first to do so. This may be a person who gets a low wage, but who is also unproductive on the farm. The assumption of Henning and Henningsen, however, is not unreasonable if off-farm returns to working time vary more than on-farm returns.

We have seen that capital and risk market imperfections can lead to a situation where separability breaks down in such a way that consumption variables affect production ones. The model as written has little to say about failures of these markets, but it is not a difficult matter to write them in (de Janvry and Sadoulet, 2004). What is important in this context is to consider whether separability is rejected because of labour market failures, which is the focus of the tests of this chapter, or whether capital and risk market failures are possible reasons for non-separability. In the tests below, an attempt is made to control for both of these potential market failures.

5.3 Reduced form separability tests

There are four possible labour market regimes – autarky, hiring in, working off-farm for a wage and both hiring in and working off farm. The sample of 980 farms with crops, summarized in Chapter 2, and used for most of the household estimations in Chapter 4, splits relatively evenly into these four groups (279, 203, 308 and 190 respectively).

The first tests of separability are at the household level for two groups – all those who work off-farm for wages and do not hire in (308 farms) and all those who hire in and do not work off-farm (203 farms). The dependent variable in both tests is total farm labour. In the first test this is equivalent to total family labour. In the second test total farm labour includes hired labour. This is the appropriate dependent variable because in this regime family farm labour is determined by consumption variables whether separability holds or not, but total farm labour is affected by consumption variables if separability does not hold (Benjamin, 1992) and not (generally) affected otherwise.

The alternative hypothesis in the first test can be summarized as non-separability owing to non-linearities in off-farm wage rewards (due, perhaps, to increasing costs of being out of the house or away from the farm for long lengths of time) or to rationing of wage labour time (perhaps due to only seasonal jobs being available[24]). The alternative hypothesis for the second test can be summarized as non-separability due either to non-linearities in the costs of hired labour (screening and monitoring costs, mainly) or to the rationing of hired labour (without the possibility of raising wages to attract more labour). By including the crop diversification variable (the inverted Simpson index), it is hoped that results are not strongly influenced by possible imperfections in the risk market. We have also seen that if there are imperfections in the capital market these are not likely to lead to most consumption variables affecting on-farm hours.

The main focus of the regressions, therefore, will be on the effects of the other consumption variables (numbers of working age adults, children and old people in the household) on the dependent variables. The multinomial sample selection model of the Appendix to Chapter 2 is re-run here in a new context, with excluded variables apparent from Table 5-A.6. Table 5-1 includes some summary statistics for the four regimes, at household level, but also at the level of the individual.[25] The information contained in this table is useful for at least two reasons. Firstly, it highlights the importance of non-agricultural self-employment (almost half as important as wage work in terms of the number of people doing it) in the sample. Secondly, it highlights the fact that even in wage earning households there are a large number of people working only on the farm or in self-employment.

Table 5-1: Household and individual work regimes[26]

	Households				Total no. of individuals
	Autarky	Hiring-in only	Wage work	Hiring-in and wage work	
Individuals					
Agriculture only	499 *(40.5)*	284 *(23.1)*	276 *(22.4)*	173 *(14.0)*	1,232 *(41.0)*
Agriculture and wage work	0 *(0.0)*	0 *(0.0)*	334 *(64.0)*	188 *(36.0)*	522 *(17.4)*
Agriculture, wage work and self-employed	0 *(0.0)*	0 *(0.0)*	24 *(61.5)*	15 *(38.5)*	39 *(1.3)*
Wage work and self-employed	0 *(0.0)*	0 *(0.0)*	20 *(74.1)*	7 *(25.9)*	27 *(0.9)*
Off-farm only	0 *(0.0)*	0 *(0.0)*	92 *(62.6)*	55 *(37.4)*	147 *(4.9)*
Self-employed only	25 *(17.4)*	43 *(29.9)*	50 *(34.7)*	26 *(18.1)*	144 *(4.8)*

[24] Robles, 1997, in cuadro 12 of his review, reports in a 1996 national survey by the Peruvian national statistics agency – INEI – where 52.9% of males among the rural poor and 34.8% from among the rural well-off do not participate in the labour force because 'there is no work'.

[25] The Table has data on 8 individual regimes: working only on the farm, working on the farm and for wages, working on the farm and for wages and in self-employment, working for wages and in self-employment only, working for wages only, working in self-employment only, not working outside the home, and working on the farm and in self-employment. 367 people in total work in self-employment, and 735 for wages. This information is used in setting up the tests of section 5.5.

[26] A multinomial regression for the 8 individual regimes of this table is reported in the appendix to this chapter in Table 5-A-3. This has a couple of interesting results on the intra-household division of labour to add to the sectoral models of the last chapter. Women in self-employment often do no work at all on the farm. The 'push' from education is towards off-farm wage work only (again – no farm work for the individual). This is true also of the Simpson crop diversification index. Farm work must be very unrewarding for these individuals to do none of it at all.

Housework only	179 *(28.2)*	141 *(22.2)*	198 *(31.2)*	116 *(18.3)*	634 *(21.1)*
Agriculture and self-employed	111 *(43.2)*	74 *(28.8)*	45 *(17.5)*	27 *(10.5)*	257 *(8.6)*
Total number of individuals in each household type	814 *(27.1)*	542 *(18.1)*	1,039 *(34.6)*	607 *(20.2)*	3,002 *(100.0)*
Number of households	280 *(28.6)*	203 *(20.7)*	308 *(31.4)*	189 *(19.3)*	980 *(100)*

To deal with this issue two versions of the reduced form tests for off-farm labour are run – in the first version only agricultural hours are included on the left hand side. In the second version (for the off-farm hours test) hours in self-employment are added to agricultural hours on the left hand side,[27] and – on the right hand side – the value of the self-employed business capital is added to farm capital (those in self-employment with paid staff are eliminated from the sample to avoid conflation of two hiring in decisions – hiring-in for the farm and for the family business). If results are unaffected by these changes, the tests using only agricultural hours are preferred. The results for the regressions are reported in Table 5–2. The multinomial participation regression (reported in Table 5–A–6) used for the sample selection version of the model in Table 5–2 is exactly the same as that used in Chapter 4. The excluded variables are farm size in 1991, years in the community, whether home owners or not, distance from a number of local amenities, an age/farm-size multiplicative variable and whether one was born locally or not. Because there may be self-selection, or market selection, according to unknown criteria (e.g. knowledge of local networks) into certain regimes, which may also affect behaviour within those regimes, it was felt that a selection approach would be informative.

[27] A separate test on the linearity of self-employed returns (reported in table 5-A-1 in the appendix to this chapter) rejects linearity, suggesting that while self-employment may possibly be conflated with agricultural work it should probably not be conflated with wage work off-farm in testing for labour market separability.

Table 5-2: Reduced form separability tests

	Separability of Hiring-In	**Separability of Working Off-Farm**	**Separability of Working Off-Farm (incl. self-emp)**	**Separability of Hiring-In: Bootstrapped standard errors (1000 replications)**
Dependent Variable	Total on-farm work	Total family on-farm work	Total family on-farm work and in self-emp.	Total on-farm work
Age	-11.512 [0.58]	-41.407*** [0.01]	-42.211** [0.02]	-15.632 [0.74]
Gender (1=male)	1,670.149** [0.04]	954.33 [0.10]	1,001.15 [0.14]	3263.923** [0.01]
Years of education	-12.344 [0.85]	-45.75 [0.19]	-25.981 [0.52]	-159.809 [0.17]
Years on-farm experience	17.977 [0.28]	63.042*** [0.00]	58.551*** [0.00]	-17.597 [0.48]
Farm tenure (higher means less secure)	-244.220** [0.04]	15.878 [0.81]	78.477 [0.29]	-321.592 [0.11]
Farm capital (+ self-emp. capital in column 3)	0.270*** [0.00]	0.380*** [0.01]	-0.012 [0.15]	0.101 [0.64]
Farm size	11.343 [0.55]	33.398* [0.06]	30.023 [0.12]	-21.218 [0.54]
Number of children (under 15)	95.036 [0.37]	-37.729 [0.52]	-60.286 [0.38]	109.339 [0.41]
Number of adult family members (15-65)	1,018.211*** [0.00]	503.146*** [0.00]	750.932*** [0.00]	881.287*** [0.00]
Number of people over 65 in the house	432.239 [0.29]	510.545* [0.06]	723.891** [0.02]	-18.442 [0.98]
Unearned income	0.15 [0.46]	0.045 [0.83]	-0.199 [0.40]	-0.21 [0.52]
Rice	-15.939 [0.98]	-430.843 [0.41]	-572.9 [0.32]	-2262.079* [0.05]
Maize	-1,473.058*** [0.00]	-312.548 [0.24]	-372.744 [0.21]	-1238.402* [0.05]
Banana	-918.002 [0.31]	-300.566 [0.51]	29.263 [0.96]	-145.035 [0.91]
Potato	-742.683 [0.18]	367.695 [0.22]	470.688 [0.17]	-1780.192** [0.02]
Coastal region	996.933* [0.08]	-308.489 [0.40]	-726.108* [0.08]	1192.747 [0.17]
Rainforest region	-88.356 [0.87]	-75.155 [0.84]	-346.953 [0.39]	557.805 [0.43]
Spanish speaking	-285.004 [0.49]	-564.257** [0.02]	-428.56 [0.14]	-320.416 [0.52]
Inverted Simpson Index	30.223 [0.86]	227.557** [0.03]	166.266 [0.15]	-160.887 [0.52]
Proportion of farm output sold on the market	457.796 [0.43]	197.794 [0.57]	344.334 [0.40]	-1705.312 [0.23]
Importance of local non-agricultural income sources	224.04 [0.84]	-2,053.208*** [0.00]	-1,844.909** [0.01]	-357.908 [0.84]
Difficulty of access (time)	-57.404 [0.35]	52.320 [0.15]	80.835* [0.07]	-38.668 [0.68]
Average distance to amenities	-2.716 [0.38]	1.675 [0.38]	1.531 [0.49]	-8.276* [0.08]
Local daily rate for agricultural labour (male)	-16.298 [0.89]			-40.876 [0.72]
Average household hourly wage for off-farm wage work		-52.035 [0.59]	-123.03 [0.27]	
Selection variable 1				5948.394 [0.13]
Selection variable 2				4191.049** [0.04]
Selection variable 3				18105.040*** [0.00]
Selection variable 4				-1556.822 [0.73]
Intercept	852.506 [0.63]	3,258.179*** [0.00]	2,932.108** [0.02]	8516.492 [0.14]
N	203	308	280	203
R^2	0.433	0.387	0.405	2033.552

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets.

The selection variables are insignificant for the off-farm test, so only OLS results are reported (with and without off-farm self-employment included). For the hiring-in test, the selection variables are significant. For the given specifications, the results suggest that separability probably does not hold, either in working off-farm or hiring in on-farm. The amount of adults in the family seems always to have a positive effect on the amount of hours worked on the farm. The results also suggest that adding self employment does not change the picture very much.

It is unknown, however, if either result is due to rationing or to non-linear transaction costs (given that these two are the most likely causes of non-separabilities).

5.4 A new test for diagnosing the reasons for non-separability.

It is possible, however, to make progress on the issue of whether rationing or non-linear transaction costs are causing non-separability. If a binding ration in off-farm work exists, then off-farm hours are clearly exogenous in a regression with on-farm work as the dependent variable, and endogenous otherwise. If, on the other hand, there are non-linearities in costs then off-farm work is endogenous (even small changes in agricultural prices or technology will affect the amount of time in work off-farm). The same kind of distinction can be made for those farms hiring in in a non-separable regime: if they are rationed, then the amount of labour hired in should be exogenous in a regression where total family labour is the dependent variable: if, on the other hand, costs are non-linear, exogeneity should be rejected.

Table 5-3 includes the results of two tests on the exogeneity of off-farm hours and hired in hours. The first test is the common Wu-Hausman test, using instruments that are listed beneath the table. The second exogeneity test is a form of the Chow test (de Luna and Johansson, 2001) based on sorting the data according to the potentially endogenous variable (in these cases, either hours hired in or hours worked off-farm) and then conducting a Chow test for parameter constancy where the sample is divided in two at the median value of the sorted variable.

Table 5-3: Testing exogeneity: instrumental variable models and chow tests

	Exog. test for off-farm wage work	Exog. test for hired-i n work
Dependent variable	Family on-farm hours	Family on-farm hours
Age	-59.326** [0.01]	-8.600 [0.70]
Gender (1=male)	688.534 [0.35]	1792.721** [0.03]
Years of education	-61.597 [0.16]	8.783 [0.91]
Years experience on-farm	80.081 [0.00]***	13.321 [0.47]
Farm tenure (higher means less secure)	38.578 [0.63]	-257.330** [0.04]
Farm capital	0.659 [0.02]**	0.333 [0.13]
Farm size	56.598 [0.05]*	16.439 [0.42]
Number of children (under 15)	-135.610 [0.21]	111.341 [0.33]

Dependent variable	Exog. test for off-farm wage work Family on-farm hours	Exog. test for hired-i n work Family on-farm hours
Number of adult family members (15-65)	211.236 [0.41]	1029.943*** [0.00]
Number of people over 65 in the house	585.462 [0.07]*	450.622 [0.34]
Unearned income	0.030 [0.91]	0.200 [0.37]
Maize	-404.682 [0.52]	-212.968 [0.69]
Potato	-379.658 [0.24]	-1722.170** [0.01]
Rice	47.093 [0.94]	-852.228 [0.36]
Banana	293.469 [0.41]	-788.895 [0.16]
Coastal region	-640.530 [0.22]	1109.972 [0.15]
Rainforest region	-325.708 [0.50]	241.693 [0.74]
Spanish speaking	-787.947** [0.02]	-326.550 [0.49]
Inverted Simpson Index	387.043** [0.03]	-11.153 [0.95]
Proportion of farm output sold on the market	324.616 [0.45]	435.892 [0.46]
Importance of local non-agricultural income sources	-1454.117* [0.09]	-29.335 [0.98]
Difficulty of access (time)	41.627 [0.35]	-34.325 [0.60]
Average distance to amenities	4.258 [0.17]	-3.521 [0.33]
Average household hourly wage for off-farm wage work	144.766 [0.47]	
Local daily rate for agricultural labour (male)		-144.610 [0.37]
Unpaid labour exchange is customary in the community		-811.506 [0.09]
Hours of off-farm wage work (household)	0.800 [0.23]	
Hours of farm labour hired in		-1.304 [0.11]
Intercept	2447.140 [0.08]	2140.337 [0.28]
N	308	203
Prob>F	0.000	0.000
Sargan statistic (overidentification test of all instruments):	3.537	5.69
$Chi^2(4)$ P-value	0.472	0.223
Tests of endogeneity – H_0: Regressor is exogenous		
Wu-Hausman F test:	3.126	4.251
F (1,281) P-value	0.08	F(1,175) 0.04
Durbin-Wu-Hausman chi-sq test:	3.389	4.814
$Chi^2(1)$ P-value	0.07	0.03
Chow test for exogeneity	Off-farm exogeneity test	Hiring-in exogeneity test
F	1.27	1.65
Crit F at 5% (10%)	1.55 (1.41)	1.58 (1.43)
	Do not reject exogeneity	Reject exogeneity

Excluded instruments for the IV regressions – Years in the community for newcomers, Blow-in, Home tenure, Farm size (1991), Total amount of years lived in cities by households. *=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets.

The results are not 100% clear cut, but are still reasonably clear. P-values are between .04 and a little higher than .18. The most likely conclusion is that separability does not hold for off-farm work because of rationing and that it does not hold for hired-in work because of non-linear costs. This latter is an important result because of what it suggests about the land distribution-equity/efficiency issue, discussed at the beginning of this section. This issue is followed up in the next chapter.

However, a potential problem with the results is that the coefficient instrumented off-farm hours is not negative, which one would expect. The instruments may be weak;

and the coefficient of the uninstrumented variable is significant negative (as seen in the appendix, where the uninstrumented regression results are reported in the second column of Table 5–A–4). The second, Chow, test, though, does not depend on the quality of instruments, and gives very close to the same result – exogeneity for working off-farm but not for hiring-in. It should also be noted that these are conditional models (conditional on participation in a market), and there may be selection effects. From Table 5.2 these seem unlikely to be important for the working off-farm test, but may have some effect on the hiring in test.

5.5 Heterogeneous households and individuals in reduced form separability testing

Testing for non-separability with reduced forms at the household level, as opposed to the individual level, may be misleading.[28] One reason why this could be so is because of the possible confounding effects of non-linearity in household wages (because the better paid may work off-farm first), even though individual wages may be linear, as noted by Henning and Henningsen (2005). About 40% of houses in the Peruvian LSMS have multiple wage earners. As seen in Table 5-1, in regime 3 (wage earners, no hiring in), there are 468 individual wage earners in 308 households. In regime 4 (wage earners and hiring in) there are 265 wage earners in 189 households. Roughly one quarter of these wage earners (112 from regime 3 and 62 from regime 4) do not do any farm work at all! It is possible in general that both the cumulative effect of extra earners[29] and the discreteness of the work choice among some household members increase rejection of household-level separability tests.

A second reason to switch to individual data is that household surveys rarely ask detailed questions of all individuals. Questions concerning rationing and transaction costs can be put into surveys (and are beginning to enter the LSMS surveys e.g. Guatemala 2000), but it is unlikely that they will be answered by everyone concerned.

[28] Even if the household is an appropriate unit of analysis, misleading cases could also occur if the sample used in the conventional reduced form test for off-farm non-separabilities includes households hiring in (conflating regimes 3 and 4). In this case, it is clear that lack of separability might be found due to non-linear costs of hiring in, irrespective of whether off-farm work is rationed or subject to transaction costs or not. Reducing the sample to eliminate households hiring in, and controlling for sample selection (as done here) should provide clearer alternative hypotheses.

[29] To allow for this case the regressions for regimes 3 and 4 of Table 2 were re-run using a sample of households with only one person working for a wage with very close to the same results, so the non-linearity of household wages does not appear to be an important factor in the rejection of separability for regimes 3 and 4. This table is in the appendix (Table 5-A-2).

So procedures for testing and measuring the degree of market imperfections will continue to be needed. And even if everybody relevant does answer the appropriate survey questions, it will still be useful to be able to confront the implications of the household model with data concerning rationing and transaction costs.

The main reason, however, to move from the household as the unit of analysis to the individual is simply to increase power by making maximum use of all available information.

Vakis *et al.* (2004) use a 'latent class' Mixture Model to separate two classes of Peruvian households who are working off farm. A version of the test regression of Table 5.2 is run, and two classes identified – one where separability holds (consumption variables do not affect on-farm family time) and one where separability does not hold (consumption variables do affect on-farm time). They find that roughly half the households are in a non-separable regime, and – from the switching regression – that the main reasons for non-separability include being far from markets, having low levels of education and not speaking Spanish. Their results are impressive and the model interesting, but there remains some doubts as to the validity of the classification. These are because of the conflation of regimes 3 and 4 in the sample, because the coefficient on wages in the separable regime is positive insignificant while it is negative in the non-separable regime (one would expect either the opposite sign arrangement or two negatives), and because crop diversification and potentially weak capital are not controlled for.

5.5.1 The econometric model

In this section, the methodology of Vakis *et al.* is adapted to apply to both household and individuals in Peru in 1994. The focus is on identifying individuals in separable and non-separable regimes. These individuals are all working for wages off-farm, and so the models are conditional models. It is unlikely, as Table 5-2 suggests, that this is a problem in itself, since sample selection appears not to be an issue, at least at the household level. Non-separability, if it is plausibly identified, is assumed to be mainly the result of rationing, rather than higher transaction costs, given the results of the last section. The words 'rationed' and being in a 'non-separable regime' are thus used interchangeably in the rest of this section.

The model is an extension of the Vakis model. By including capital, unearned income, main crops and diversification information, the hope is that sharper results will emerge in these specifications. The re-specification also extends to running the model at an individual level – although the household level model will also be estimated to compare results.

Following Vakis *et al.* the model estimated is as follows:

$$l^1 = x_1\beta + u_1 \qquad (5-7)$$

$$l^2 = x_2\gamma + u_1 \qquad (5-8)$$

$$\lambda^* = x_\lambda\xi + u_\lambda \qquad (5-9)$$

The first two vectors of x variables include the variables introduced in the econometrics of Chapter 4. The subscripts on x refer to regime – 1 indicating a non-separable regime and 2 a separable one. The third set of x variables are those variables that are likely to affect the possibility of being rationed.

β, γ *and* ξ are coefficients, and the three errors are normally distributed, each with a constant variance (the variance of the third is scaled at 1). l^1, l^2 *and* λ^* are latent variables. We actually observe:

$$l^i = \begin{Bmatrix} l^1 & if & \lambda^* < 0 \\ l^2 & if & \lambda^* \geq 0 \end{Bmatrix}. \qquad (5-10)$$

Given that we cannot identify a priori which regime a person is in, a randomly selected observation l_k^i (on farm hours supplied in regime i) will have a probability 1-$\lambda = \Phi(-x_{k\lambda}\xi)$ of being in regime 1 and a probability of λ of being in regime 2. The density function for person k, therefore, is:

$$f(l_k^i) = (1-\lambda)\varphi_1(l_k^i - x_{k1}\beta) + \lambda\varphi_2(l_k^i - x_{k2}\gamma) \qquad (5-11)$$

and the likelihood function is:

$$L(\beta,\gamma,\xi,\sigma_1,\sigma_2) = \prod_{k=1}^{N} f(l_k^i) \qquad (5-12)$$

This likelihood function may be solved using the normal algorithms (Greene, 2004), but Vakis *et al.* use the E-M algorithm (Hartley, 1978), after having chosen starting values from an OLS regression on the x variables from regime 1. In this section, the E-M algorithm is also used. Estimates for λ are used to weight the observations and separate regressions are run for regime 1 and regime 2. To help identification of the two classes, the hours worked for wages variable is omitted for the regime 2 estimation (as noted in section 5.4, hours worked off-farm should not affect on-farm supply of the individual in an unconstrained regime, but should in a constrained regime if the constraint takes the form of a ration).

In analysing the results of the labour supply regressions in the two regimes, it is important to note that, unlike in the household models, the family or household variables (number of children, working age family members, older family members) are not especially informative in these individual models. These variables affect the labour of all family members working only on-the farm or in self-employment in ways that can lead to more or less on-farm labour for each individual, and hence to more or less on-farm labour by any unconstrained off-farm wage worker. In the individual models below there are a number of other coefficients to focus on to help decide if the latent class division, along the lines suggested by the model, is successful. For example, one would expect the results to show a negative effect of wages on hours worked in the unconstrained regime (because of a substitution effect away from farming) and a probably weaker effect in the constrained regime (only the income effect is relevant in this case).[30] One would also expect the farm labour productivity variables (farm size, efficiency proxies and farm capital) to come through stronger in a positive direction in the unconstrained regime than in the constrained regime because there is no income effect on on-farm work in the unconstrained case (the income effect changes the amount of off-farm work instead).

The results of the third estimation (the 'switching' estimation, where λ is estimated) are very sensitive to the chosen x variables. It is often difficult upon completion to know whether one is at a global or local optimum, and likelihood values need to be compared for a number of model versions. The final version chosen is quite

[30] To work out the precise effects on each person's farm labour time in either regime resulting from a change in off-farm wages of one person is a reasonably straightforward comparative static problem (Lee, 1998, ch. 6). The working assumption here is that the labour of each family member is a gross substitute with that of other family members, and that leisure time is also a (Hicksian) substitute and is normal. These are not unrealistic assumptions in poorer countries.

parsimonious in the switching equation – education, being a non-main person, gender, on-farm experience, farm size and the regional and broad sectoral dummies were the variables used. These variables are not highly correlated with each other. Using these variables the sample is neatly divided along the grounds expected by theory (into apparently rationed and non-rationed groups), with the results that 43% of the sample are placed as rationed, and 57% unrationed.

Tables 5.5 and 5.6 contain, respectively, the estimation results and summary statistics for the designated classes (i.e. those with an unconditional expectation of being in the respective class). The switcher estimation suggests that the main people rationed in off-farm work will be operators with low education, on small farms, with slightly less experience than others, living in the sierra. These are the same results, except for the farm experience variable, as are got for the same equation at household level, in the appendix (in tables 5-A-4 and 5-A-5).

The summary statistics of Table 5-6 provide more information on those in the two designated classes. The unconstrained majority have a higher per capita income than the others; they work more off-farm and less on-farm; they have higher wages; they travel further to work and are more likely to do seasonal work; they do not work in construction, especially not in skilled construction labour, and they are slightly less likely to work off-farm than in agriculture and a good deal more likely to work in the professions; they have bigger better capitalized farms; they have far higher levels of education and they are far less likely to be Spanish speaking. Thus, it would seem the constraints become binding when assets and education are low, and are merely theoretical otherwise.

Table 5-4: Heterogeneous individuals in constrained and unconstrained classes

Dependent variable	Probability of being constrained	On-farm hours	On-farm hours (for the constrained)	On-farm hours (for the unconstrained)
Off-farm individual wage		-138.971* [0.09]	-169.964*** [0.01]	-159.686*** [0.00]
Average education of others in the house		-15.104 [0.35]	4.305 [0.77]	9.079 [0.35]
Average gender of others in the house		-29.907 [0.88]	-739.618*** [0.00]	379.904*** [0.00]
Average age of other adults in the house		0.483 [0.95]	17.571*** [0.00]	1.157 [0.78]
Average on-farm experience of others in the house		-9.420* [0.08]	-27.099*** [0.00]	-11.761*** [0.00]
Average off-farm experience of others in the house		0.312 [0.97]	4.760 [0.45]	7.147 [0.15]
Age		16.523 [0.46]	10.175 [0.51]	19.211* [0.10]
Age^2		-0.218 [0.38]	-0.162 [0.34]	-0.161 [0.24]
Gender (1=male)	0.511*** [0.00]	443.545*** [0.01]	31.233 [0.76]	351.478*** [0.00]
Non-main person	-0.970***	-277.79	335.194**	9.955

Dependent variable	Probability of being constrained	On-farm hours	On-farm hours (for the constrained)	On-farm hours (for the unconstrained)
	[0.00]	[0.20]	[0.02]	[0.93]
Years of education	-0.086*** [0.00]	-6.721 [0.65]	44.456*** [0.00]	22.403** [0.02]
Years on-farm experience	-0.010*** [0.00]	11.308*** [0.01]	28.596*** [0.00]	8.129*** [0.00]
Years experience off-farm		-0.790 [0.88]	-2.635 [0.45]	-4.753* [0.10]
Farm tenure (higher means less secure)		28.654 [0.35]	-8.214 [0.72]	11.574 [0.55]
Farm capital		0.054 [0.43]	0.249*** [0.00]	0.053 [0.14]
Farm size	-0.021*** [0.00]	6.677 [0.45]	7.783 [0.65]	25.436*** [0.00]
Number of children (under 15)		-6.713 [0.80]	-30.933 [0.14]	37.228* [0.06]
Number of adult family members (15-65)		-9.497 [0.83]	-9.510 [0.72]	-6.120 [0.83]
Number of people over 65 in the house		-161.967 [0.29]	-700.408*** [0.00]	-2.304 [0.98]
Unearned income		0.047 [0.70]	0.166* [0.10]	0.010 [0.89]
Inverted Simpson Index		-213.159 [0.35]	-421.318** [0.04]	-454.496*** [0.00]
Rice		-134.806 [0.27]	-255.765*** [0.01]	-158.781** [0.03]
Maize		-95.18 [0.67]	-689.729*** [0.00]	163.862 [0.12]
Banana		169.424 [0.23]	-324.425*** [0.00]	19.930 [0.85]
Potato	-1.248*** [0.00]	-185.798 [0.30]	94.001 [0.56]	-119.167 [0.20]
Coastal region	-0.700*** [0.00]	-113.924 [0.49]	594.206*** [0.00]	-168.230* [0.08]
Rainforest region		-69.753 [0.56]	-328.932*** [0.00]	59.998 [0.39]
Spanish speaking		117.102** [0.02]	68.307** [0.04]	17.778 [0.61]
Proportion of farm output sold on the market		283.364 [0.10]	96.099 [0.52]	376.022*** [0.00]
Importance of local non-agricultural income sources		-688.713 [0.02]	-3109.356*** [0.00]	225.094 [0.35]
Difficulty of access (time)		37.201** [0.05]	20.621 [0.17]	-1.240 [0.92]
Average distance to amenities		0.487 [0.60]	-0.219 [0.75]	-0.194 [0.69]
Household hours of wage work		-0.304*** [0.00]	-0.432*** [0.00]	(dropped)
Manufacturing Sector	0.090 [0.55]			
Farming Sector	0.133 [0.33]			
Services Sector	-0.046 [0.79]			
Intercept	0.630*** [0.00]	1596.177*** [0.01]	4807.821*** [0.00]	-705.411 [0.09]
Number of obs	317	317	317	317
R^2	0.572	0.327	0.7772	0.43
Adjusted R^2	0.558	0.248		

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets.

Table 5-5: Summary statistics for the constrained and the unconstrained

	Constrained Group			Unconstrained Group		
Variable	N	Mean	Std.Dev.	N	Mean	Std.Dev.
Income per capita	137	882.28	715.38	180	970.46	869.34
Hours in wage work	137	946.04	794.50	180	1003.10	862.78
Hours in farm work	137	1616.41	952.52	180	1210.46	842.61
Wage earnings from manufacturing	137	209.91	523.17	180	154.74	533.31
Wage earnings from the service sector	137	42.08	262.84	180	245.51	898.38
Earnings from agricultural wage labour	137	457.15	591.33	180	525.10	673.02
Average household hourly wage for off-farm wage work	137	0.76	0.44	180	0.93	0.69
Average education of others in the house	137	5.77	3.58	180	6.97	3.24
Average gender of others in the house	137	0.31	0.34	180	0.42	0.35
Average age of other adults in the house	137	32.03	9.92	180	34.17	11.42
Average on-farm experience of others in the house	137	11.02	10.40	180	11.20	11.81
Average off-farm experience of others in the house	137	2.00	4.55	180	4.33	6.99
Incomer to the village (all)	137	0.15	0.35	180	0.29	0.45
Age	137	43.60	14.29	180	34.93	13.71

	Constrained Group			Unconstrained Group		
Variable	N	Mean	Std.Dev.	N	Mean	Std.Dev.
Gender (1=male)	137	0.88	0.32	180	0.77	0.42
Non-main person	137	0.02	0.15	180	0.33	0.47
Years of education	137	5.14	3.64	180	8.15	3.89
Number of rooms in the family home	137	2.37	1.06	180	2.67	1.35
Hours of farm labour hired in	137	0.00	0.00	180	0.00	0.00
Years on-farm experience	137	20.63	14.84	180	18.13	14.05
Years experience off-farm	137	7.65	10.71	180	7.79	10.82
Off-farm individual work insured	137	0.04	0.21	180	0.07	0.25
Off-farm individual work in public sector	137	0.21	0.41	180	0.11	0.31
Cost of individual eating out	137	0.07	0.50	180	0.06	0.58
Farm capital	137	654.37	792.66	180	433.61	670.42
Farm size	137	1.53	1.80	180	3.66	7.52
Number of children (under 7)	137	1.43	1.29	180	1.39	1.26
Number of children (under 15)	137	2.48	1.85	180	2.73	2.08
Number of people over 65 in the house	137	0.15	0.42	180	0.13	0.41
Number of adult family members (15-65)	137	2.98	1.27	180	3.43	1.57
Inverted Simpson Index	137	2.38	1.03	180	2.08	1.08
Proportion of farm output sold on the market	137	0.32	0.33	180	0.42	0.37
Unearned income	137	132.82	386.23	180	180.37	417.62
Rent from property	137	159.15	205.95	180	332.75	463.65
Transfers	137	126.77	404.76	180	241.59	827.81
Importance of local non-agricultural income sources	137	1.03	0.12	180	1.05	0.19
Difficulty of access (distance)	137	2.59	0.96	180	2.59	0.94
Difficulty of access (time)	137	2.02	2.89	180	1.19	2.32
Farm tenure (higher means less secure)	137	1.52	1.19	180	2.28	1.88
Coastal region	137	0.00	0.00	180	0.26	0.44
Rainforest region	137	0.08	0.27	180	0.35	0.48
Spanish speaking	137	0.30	0.46	180	0.62	0.49
Average Time to Get to Work (minutes)	137	9.83	23.28	180	14.85	26.59
Seasonal Labour (1=seasonal)	137	0.08	0.27	180	0.20	0.40
Construction		0.3	0.19		0.04	0.20
Agriculture		0.66	0.48		0.60	0.49

How likely is it that the sample has been correctly placed into constrained and unconstrained groups? Obviously, some sort of clustering has taken place, and the procedure has identified a fairly well paid cohort of about 57% or so of wage workers in the sample. The following aspects of the results suggest that the identification procedure probably has worked, although like all forms of cluster analysis it is no more than an exploratory procedure, and one must allow for some fuzziness in the separation process.

As expected, the wage coefficient in the unconstrained estimation of Table 5-5 is significant negative (as it is in the other regression also). Also as expected, the farm size and market embeddedness variables are positive significant for the unconstrained weightings, but insignificant for the other class. These variables are strong indicators of efficiency on the farm (as shall be seen in the next chapter) and a higher marginal product of labour. The capital variable, however, is stronger in the constrained group (possibly because of the amount of livestock embodied in the capital variable: higher capital may not necessarily indicate a large increase in labour productivity). The demand variables are not significant in the unconstrained regression, which is more likely if constraints do not bind. However, the summary statistics probably give as clear or clearer evidence than the regressions that a real division has been carried out by the algorithm, using this individual data[31].

5.6 Conclusions

In this chapter a new test for off-farm labour rationing has been proposed, and a recently developed 'latent class' procedure for exploring labour market separability has been tested for only the second time. The latent class procedure, which up to now has only been used once on a sample using household data, is used here on individual data (which, at least in this case, seem to give clearer results than in the household version).

The results of reduced form separability tests at the household level suggest, initially, that separability in working off the farm does not hold. It is not generally known if this is due to non-linear transaction costs, or non-linear wages (it is not clear in the literature very often how the self-employed are treated in estimations), or to rationing. The testing procedure generally suggests that the self-employed do have non-linear wages and so should not be included on the right hand side in reduced form separability tests, as seems to be quite common. Also, the multiple earner non-linear wages hypothesis of Henning and Henningsen (2005) is rejected: the separability results for households with one earner do not differ from those with multiple earners. With regard to separability itself, the results of the new test, intended to test for the cause of non-separabilities, suggests that rationing of the off-farm labour market probably does exist. The implementation of the 'latent class' exploratory procedure then suggests that those most affected by rationing-type constraints tend to be people in households with lower overall income, probably on less profitable farms and with less education – generally, people who need off-farm work more. Having off-farm experience oneself, or having a family with off-farm experience, tends to be associated with having weaker constraints.

The chapter also showed that there appears to be no rationing of labour hired in, but non-linear costs of hiring in appear to exist (the traditional explanation for this finding

---

[31] The results for the household version of the model are relegated to the Appendix in tables 5-A-4 and 5-A-5. While substantially the same two classes seem to be identified as in the individual level regression, about 76% in this regression are deemed unrationed, and the coefficient on wages is not significant negative as one might expect (while it is in the rationed group). The summary statistics are also quite similar, but there is a probability, from the wages coefficient result in particular, that the individual regression has provided the sharper clustering.

in the literature is rising monitoring costs). The amount of hiring-in generally is low, and is especially low in areas where minka or chova-chova or other customary labour exchange activities are common. Non-linear transaction costs on hired labour have been used (along with the existence of some other market imperfection, such as an imperfect land market) to justify, on efficiency grounds, a roughly equal farm size distribution, where farms are of a size that can be most productively farmed by the farm family, without hiring in (Carter and Wiebe, 1990; Eswaran and Kotwal, 1986; Feder, 1985; Sen, 1966; Swamy, 1981; Wiens, 1977). The findings of this chapter suggest that for the farms in rural Peru equity and efficiency probably are aligned. This last finding receives support in the next chapter.

6

Efficiency, permanent Income and off-farm work

Introduction

Although much of the focus of this thesis is on intra-household issues, this chapter is used to test the importance of off-farm work for households as a whole over the long term. To this end two sets of regressions are run. In the first, the relationship between off-farm work (whether measured in hours, or number of income sources) and a variety of efficiency measures is estimated. In the second, the relationship between off-farm work and permanent household income (as measured by total household expenditure) is estimated. The main objective of the chapter is to contribute to the research stream on household income strategies (Reardon *et al.* 2001). This is a growing area of research but the effects on efficiency and on permanent income over the long term of differing strategies have rarely been examined.

6.1 Off-farm work and efficiency

In this part of the chapter, the relationship at the household level between off-farm work and technical and allocative efficiency is examined. There are several reasons why off-farm work should affect both types of efficiency (and indeed scale efficiency also). Firstly, technical efficiency may be improved by know-how gained and contacts made in off-farm work. It may be improved if improved cash-flow due to off-farm work results, through a smoother flow of working capital, in a more efficient use of factors or better quality factors from more distant markets. On the other hand it may be if off-farm work reduces managerial capacities on the farm, leading to a more extensive type of farming. van der Ploeg (1990) argues that this is the case for one part of the Peruvian sierra.

Secondly, allocative efficiency may also be affected by off-farm work. If off-farm work is risk reducing, then allocative efficiency on the output side may improve if less risky but low-priced crops are being produced to minimize risk. If apparent allocative inefficiency is not really an inefficiency at all, but a deliberate strategy necessitated by the failure of local food markets, say, then it is quite likely that off-farm work will have no effect on measured allocative efficiency whatsoever.

Endogeneity in efficiency models with off-farm hours/earnings on the right hand side is certainly possible. This is so because similar household characteristics could be related both to on-farm efficiency and to off-farm behaviour. To attempt to control for this problem off-farm hours/earnings are instrumented in all the efficiency regressions of this chapter.

A by-product of the study of the relation between off-farm work and efficiency is that it also allows the discussion of the relation between education and farm efficiency. This issue remains a very contentious one (Fafchamps and Quisumbing, 1997). A survey by Lockheed *et al.* (1980) contains evidence on 39 equations from 18 studies in 13 countries and concludes that, on the whole, education increases farm productivity. Phillips (1987), however, argues that these results vary substantially by economic region. Studies from Asia support a positive significant effect of education on on-farm productivity, but evidence from Latin America and Africa is mixed. We have seen that education increases the probability of entry into well-paid off-farm farm work and that higher education is associated with higher off-farm earnings and overall household income. The relationship between education and farm productivity is something that has not yet been looked at.

6.1.1 Data envelopment analysis and stochastic frontiers

To measure efficiency, non-parametric Data Envelopment Analysis (DEA) is used. Economists tend more often to use a parametric approach rather than a non-parametric approach to measure efficiency on farms. The main parametric approach used is stochastic frontier estimation (Coelli *et al.* 1998). One disadvantage of DEA as far as economists are concerned is that it is not easily amenable to statistical analysis. Only very recently has this been possible, mainly through the use of bootstrapping techniques. These have been used in three main ways – to calculate standard errors and estimate biases for efficiency measures (Simar and Wilson, 2000), to make adjustments due to differences in sample size across regions (Zhang and Bartels, 1998; Frazier and Graham, 2005) and to deal with the inherent dependency of DEA efficiency estimates (Xue and Harker, 1999) in a regression context.

Since one of the aims of this chapter is to focus on the relationship between household livelihood strategies and different kinds of efficiency measures, bootstrapping all the estimations is not feasible (a future task might be to do so for one particular efficiency measure). Furthermore, the post-DEA regression is complicated by both censoring and endogeneity, making bootstrapping for the reasons Xue and Harker suggest (the lack of independence of DEA measures, due to the fact that they are essentially relative efficiency measures) extremely difficult. The censoring results from the fact that all measures are in the [0,1] interval. The endogeneity results from having off-farm labour (either in the form of hours or number of income sources) on the right hand side of the regressions. It is as likely that either measured technical efficiency or allocative efficiency affects income strategies in the household as it is that income strategies affect technical or allocative efficiencies. It is quite possible also that missing variables affect both. So endogeneity is an extremely likely problem.

One other disadvantage of DEA compared to stochastic frontier analysis is that it is often felt to be sensitive to outliers or errors in the data (Coelli, 1998). The LSMS data is fairly reliable data on a national level (Deaton, 1998). DEA can also be sensitive to sample size (a larger sample size reduces the proportion of farms achieving efficiency) and to the degree of input and output aggregation.

However, Kalirajan and Shand (1999) show that DEA outperforms stochastic frontier analysis if the underlying technology is unknown (that is, if the chosen functional form does not happen to accidentally fit the unknown technology), which is certainly the case here.

One other attraction for economists of the programming approach (DEA) (apart from the robustness to functional form issues) is that it allows easily for multiple outputs.

The non-parametric approach to measuring efficiency was first formulated as a linear programming model by Charnes *et al.* (1978), following on Farrell's 1957 posing of the question of relative technical efficiency in the form of a unit isoquant model. DEA has been used to measure technical efficiency in agriculture relatively frequently (Weersink *et al.* 1990; Cloutier *et al.* 1993; Chavas *et al.* 1993; Ray *et al.* 1993; Townsend *et al.* 1998; Jaforullah *et al.* 1999; Sharam *et al.* 1999; Fraser *et al.* 1999; Zaibet and Dharmapala, 1999; Jha et al., 2000; Shafiq et. al, 2000; Brummer, 2001;

Fraser *et al.* 2005; Chavas *et al.* 2005), but only the third, fourth, tenth and last named go on to measure allocative efficiency.

Before proceeding, it is important to clarify that it is only the DEA technical efficiency measure that can in any way be regarded as a true efficiency measure. The other two measures (allocative and scale efficiency) are almost certainly better regarded as measures of the responses to constraints imposed on individual farms by their exposure to market failures. What may seem like an allocatively inappropriate mix of activities will usually be undertaken because of insurance market failures or consumption goods market failures. What may seem like a scale inefficiency (e.g. farms being 'too big' for the most efficient use of inputs) may also be a response to insurance market failures (e.g. scale efficiencies could be due to something like an apparent overstocking of inputs such as traction animals, which are indeed inputs but may also be useful assets in case of emergency). While the terms allocative inefficiency and scale inefficiency will continue to be used, they are regarded as mere labels for complex phenomena which may not have any inefficiency aspect to them.

6.1.2 DEA models

There is no data on input prices, so output oriented efficiency measures only are calculated (this allows a convenient agnosticism about how perfect input markets are: from the last chapter it seems that the market for hired labour may well be imperfect).

The linear programme for the basic output oriented DEA is:

$$\max_{u,v} \sum_k u_k y_{k,j0}$$

$$\text{s.t.} \sum_i v_i x_{i,j0} = 1$$

$$\sum_k u_k y_{k,j} \leq \sum_i v_i, x_{i,j} \forall j$$

$$u_k, v_i \geq 0 \qquad (6-1)$$

y are the outputs, indexed by k. x are the inputs, indexed by i. Each farm is indexed by j, and u and v are output and input weights. This programme has to be solved for each farm (each $j0$), where the weighted output for each farm is maximized with the proviso that the efficiencies of all other farms is less than or equal to 1. This gives us a relative measure of efficiency, where efficiency is defined as weighted output divided by weighted input. The denominator – weighted input – as we can see from the first constraint – will be equal to 1. In DEA each farm optimally sets its own weights. Hence the need for separate farm by farm solution of the problem.

The farm level measure of efficiency is the value of the objective function. It measures constant returns to scale (CRS) technical efficiency. If variable returns (VRS) are allowed, and scale efficiency and technical efficiency are to be analytically separated, then a further constraint is added: the sum of the output weights must be equal to 1. This ensures that farms are being compared for technical efficiency with other farms of similar size, so the data envelope fits closer, and (pure) technical efficiency measures are higher (or the same). Scale efficiency measures can be found by dividing pure technical efficiency by CRS technical efficiency measures. The programme can then be estimated for a third time, where the VRS constraint is changed so that the sum of output weights is less than or equal to 1. In this case farms are being compared with other farms of the same size or smaller. If this efficiency measure is the same as the pure VRS measure, then diminishing returns exist (the farm is 'too big'), and if it is not the same, then increasing returns exist (the farm is 'too small'). Allocative efficiency measures can be found by changing the objective function to one where revenue is maximised, subject to the same constraints, where measured output is first adjusted to account for technical inefficiencies (all outputs are divided by the farm level technical efficiency coefficient). Hypothetical revenue (were output on the technical frontier) is then divided by maximum possible revenue (the new objective function) to yield, again, a measure of efficiency in the [0,1] interval.

6.2 Data and variables

The output oriented DEA model has been separately estimated for Peru for three regions – the sierra, the coastal region and the rain forest – based on the assumption that each region has a different underlying technology. The 14 outputs chosen are rice, potatoes, maize, bananas, cebada grano, quinua, trigo, arveja, habas (a legumbre), habas (a menestre), oca, maiz chala, maiz duro and frijol (a legumbre). These are the 14 most commonly produced crops. Not all are produced in each region and an output is eliminated from the regional model if less than 5% of farms produce it (3 outputs are eliminated from the sierra, 1 from the coast and 2 from the jungle regions; for the few farms concerned they tended to be minor outputs). For 685[32] farms at a national level, at least 90% of their crop output comes from among these 14 crops, and this is the criterion for selection to the sample. Livestock outputs are not included, so measures will be biased to some extent. These outputs are difficult to measure. Traction livestock, however, is important as a crop input and is included as part of the measure of farm capital (see Appendix 2, Chapter 4). Livestock specific inputs (livestock input costs plus hired labour for livestock) are excluded, to reduce the bias from leaving out livestock outputs. The five inputs included are farm size, capital, hired labour cost for crops, cost of crop inputs (insecticide, fertilizer etc.) and family farm labour.

Self-declared price information is also available in the survey (median segmento prices are used), so allocative efficiency measures are possible.

6.3 Efficiency results

The focus in this section is mainly on the sierra, where around 65% of the sample come from. While summary statistics for the three efficiency measures are given in Table 1 for all regions, only the sierra figures are used later in the estimations. It is important not to read too much into the inter-regional differences, because the difference in sample sizes is likely to exaggerate efficiency for the coastal region and diminish it for the sierra. Nevertheless, the differences are quite stark. The intra-regional comparisons of the three efficiency measures are also interesting:

Table 6-1: Technical, Allocative and Scale Efficiencies in the Three Regions

Region		N	Mean	Std. Deviation	Minimum	Maximum
Sierra	Technical Efficiency (VRS)	378	0.54	0.37	0.00	1
	Allocative Efficiency	378	0.45	0.28	0.00	1
	Decreasing Returns to Scale	378	0.55	0.50	0.00	1
	Increasing Returns to Scale	378	0.11	0.32	0.00	1
	Scale Efficiency	378	0.83	0.25	0.09	1
Coast	Technical Efficiency (VRS)	63	0.82	0.27	0.00	1
	Allocative Efficiency	63	0.74	0.30	0.00	1
	Decreasing Returns to Scale	63	0.13	0.34	0.00	1
	Increasing Returns to Scale	63	0.30	0.46	0.00	1
	Scale Efficiency	63	0.91	0.19	0.17	1
Rainforest	Technical Efficiency (VRS)	164	0.56	0.32	0.00	1
	Allocative Efficiency	164	0.38	0.28	0.03	1
	Decreasing Returns to Scale	164	0.24	0.43	0.00	1
	Increasing Returns to Scale	164	0.52	0.50	0.00	1
	Scale Efficiency	164	0.76	0.27	0.07	1

[32] This is reduced in the reported results to 605 for the estimations in order to keep this sample within the group of 980 households used in the econometrics throughout (almost) the whole thesis.

In general, allocative efficiency is lower than technical efficiency which is lower, in turn, than scale efficiency in all three regions. That allocative efficiency is low is most likely due to the diversification that takes place in all farms, but especially in the sierra (where maize or wheat are often produced for home consumption, and potatoes, among other crops, for the market). From Table 2-9, it is notable that diversification is second highest in the rainforest and lowest in the coast. It is notable also that many farms in the sierra are 'too big' in the sense that halving inputs would lead to less than a halving of outputs: more farms with less total inputs on each would be more productive. Farm size is already small in the sierra (a median of one hectare). Possibly leaving out livestock output has contributed to this result. Perhaps also it is indicative of the lack of other opportunities in the sierra; from Table 2-9 we see that family on-farm hours are higher in the sierra than elsewhere, despite the smaller farms, and so certain factors, especially family labour, are overused there. It may also be, as already suggested, that livestock are kept as security and their pure capital value as traction animals for crop production is overestimated for the sierra.

Before attempting to model the relationship in the sierra between off-farm work and efficiency measures, a brief visual inspection of the relationships between other variables and efficiency will be carried out for the sierra only. The first of the variables of interest, in Figure 6-1 below, is the 'proportion of crops sold', or the market embeddedness variable. One would expect more commercial farmers to be more efficient, and this is exactly what we find. There is a strong relationship between market embeddedness and efficiency, especially for allocative and (even more for) technical efficiency.

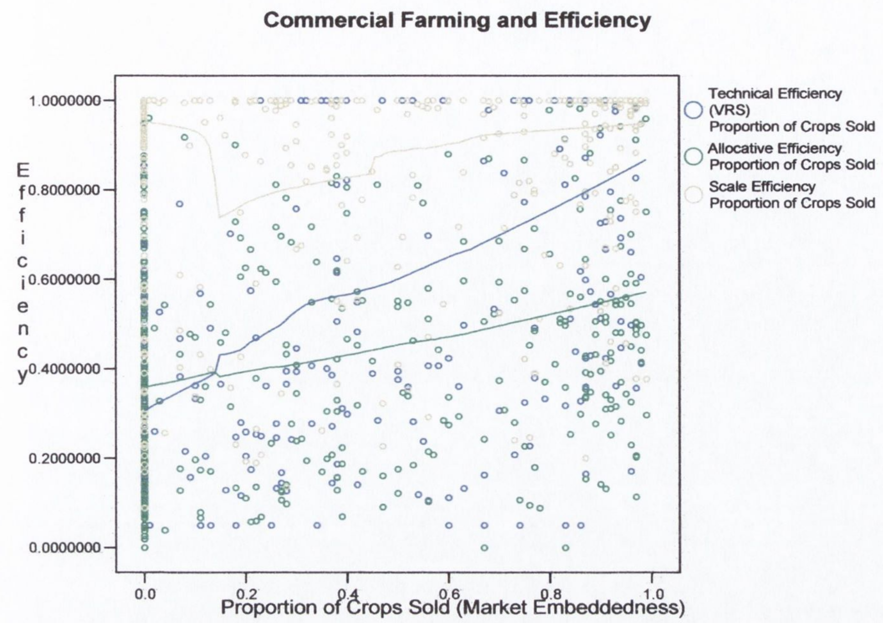


Figure 6-1: Proportion of crops sold and efficiency (with LOESS lines drawn in).

The second variable of interest is on-farm diversification (the inverse Simpson index). This measure, as used here, is higher the greater number of crops there are on a farm and the more evenly distributed is the (actual plus imputed) revenue from them. A high value is interpreted as a low cost of diversification. From the figure it is clear that this variable has a negative relationship with allocative efficiency, but a positive relationship with technical efficiency. The first relationship is not a surprise. The second, perhaps, is. But most farms in the sierra, as already noted, have plots that are far from each other, at different altitudes, suitable for different crops and it is possible that yields are higher on small, separated plots, than on larger plots, with varying quality. It may be that the skills and attention of farmers with multiple plots are greater than those with fewer plots. It is conceivable also that diversification, here, is a proxy for farm size and that Figure 6-2 is really a picture about the relationship between size and efficiency. But the pairwise correlation between farm size and diversification is just .168 for the sierra as a whole, and we shall see below in the multivariate analysis that the diversification variable tends to be far more important in explaining efficiency differences than the farm size variable. A positive relationship between diversification and technical efficiency does appear to exist.

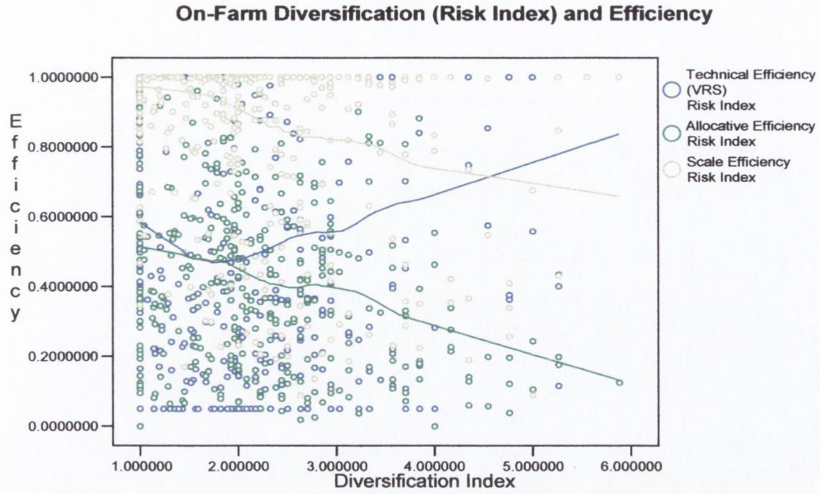
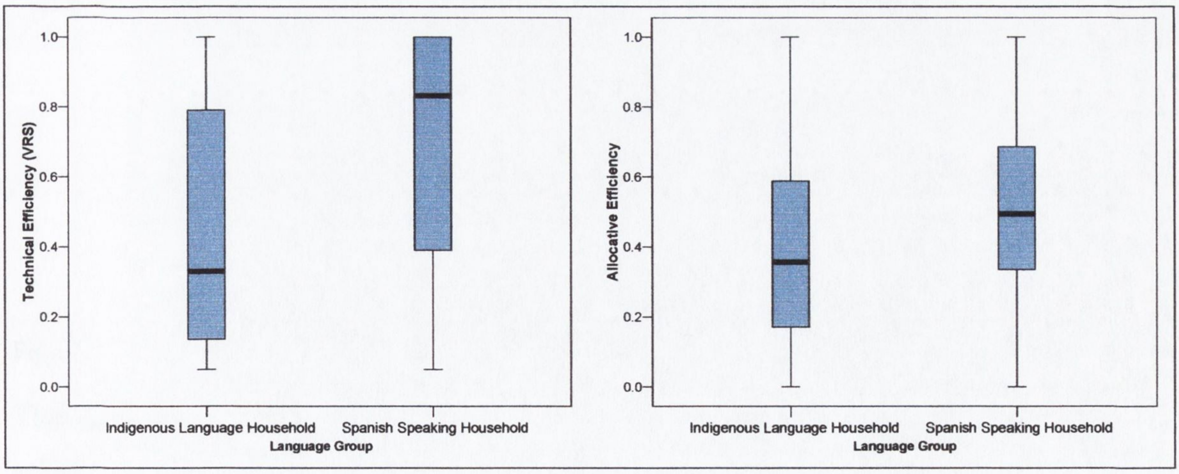


Figure 6-2: Risk index and efficiency (with LOESS lines drawn in)

The three box plots in Figure 6-3 show a strong relationship between all forms of efficiency and language group. xxxIn each case in these box plots the values of an efficiency variable are plotted on the vertical axis and the two language categories are plotted on the horizontal axis. These plots display the median efficiency level as well as the full four quartiles for each of the language groups. They also display a number of outliersxxx. van der Ploeg (1990), among others, has spoken of the farm labour sharing practices among indigenous groups, and the fact that these may reduce efficiency (there are instances in his study of farmers pulling out of such practices precisely to improve efficiency).



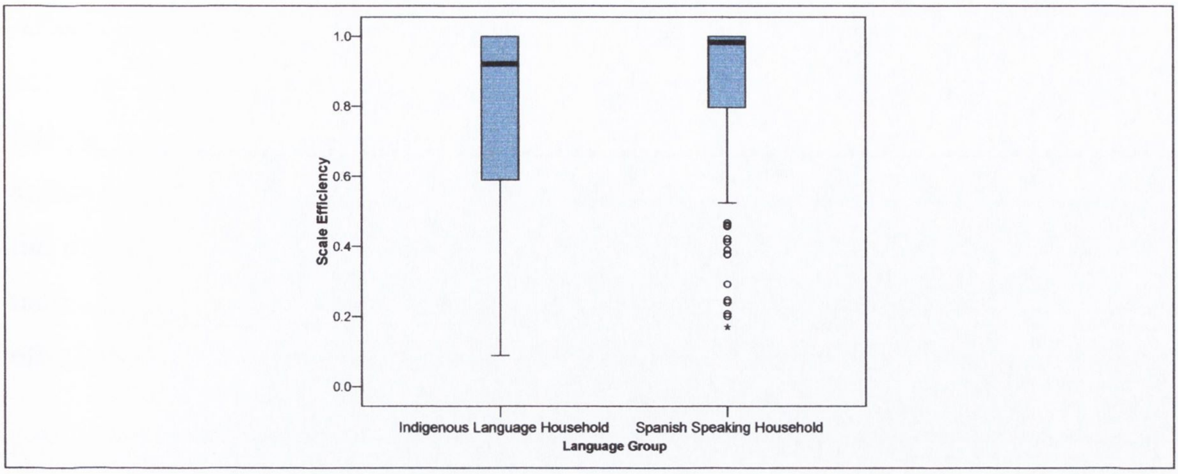


Figure 6-3: Language Group and Efficiency

The relationship between language group and efficiency is far stronger than that between education and efficiency (which only appears to turn positive when the household average schooling level reaches 8 years or so), as seen in Figure 6-4. The tangled relationship between education and on-farm efficiency was mentioned at the beginning of this chapter, and Figure 6-4 provides few answers. Certainly, the relationship is not a strong one, and it appears not to be linear.

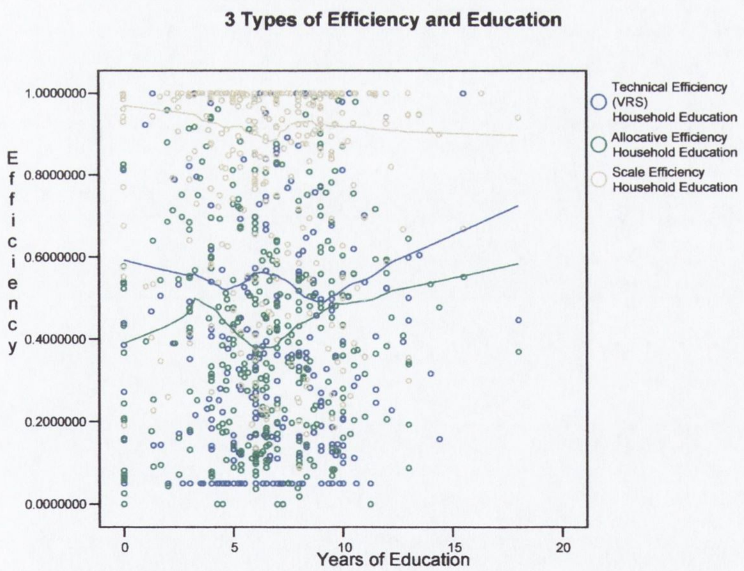


Figure 6-4: Education and efficiency (with LOESS lines drawn in)

The final two visual summaries are of the relationships between farm size and efficiency and between hours of off-farm work and efficiency. In Figure 6-5 we see that the relationship between size and efficiency is highly non-linear. Technical efficiency increases from .5 hectares up to 2 hectares (smaller farms have been omitted from this picture for clarity, as have the few large farms above 8.2 hectares).

After 2 hectares the relationship weakens. Technical efficiency declines for a while, and then rises for the largest farms. At the same time, however, scale efficiency is falling. To judge their combined effect CRS efficiency is also overlaid on the scattergram. This declines also, giving weight to the argument in the literature, and to the evidence of the last chapter concerning transaction costs for hired-in labour, that increasing farm size beyond the size suitable for a family farm reduces overall efficiency.

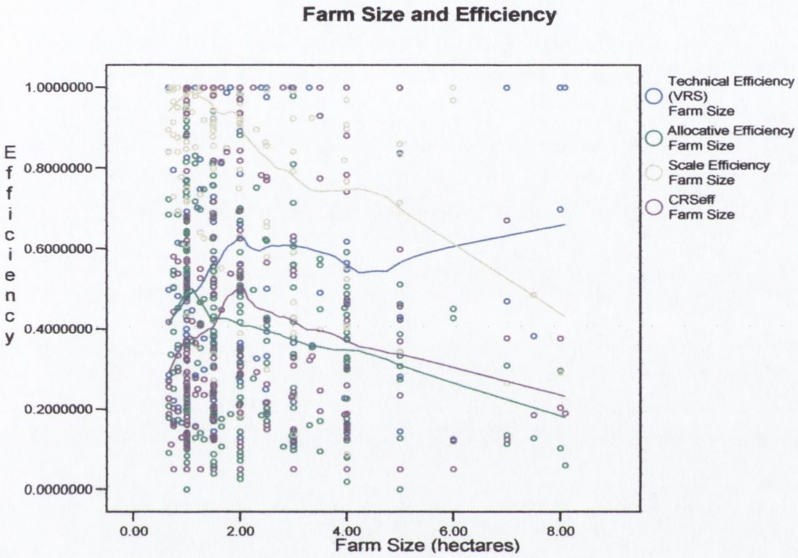


Figure 6-5: Farm size and efficiency (with LOESS lines drawn in)

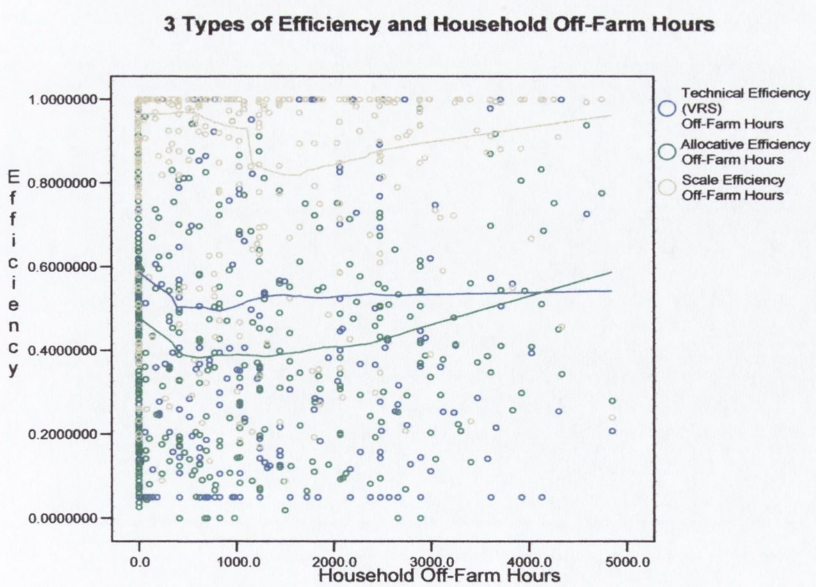


Figure 6-6: Hours off-farm and efficiency (with LOESS lines drawn in)

In Figure 6-6 the most notable feature is probably the positive relationship (after about 1500 hours or so) between off-farm work and allocative efficiency. This is not implausible if off-farm work partially substitutes for on-farm diversification as a method of risk management (the two variables have a weak, but significant, correlation of -.169 for the sierra). After initial dips for people working a small number of hours off-farm, there also appear to be positive relationships between off-farm hours and technical efficiency (weak) and scale efficiency (quite strong).

6.4 Econometric estimations

The econometric models of this section are run only for the Peruvian sierra. The dependent variable in each case is an efficiency measure, and the independent variables are limited to the six variables that were focused on individually in the last section, plus a gender variable (household average gender), since gender is often a focus in efficiency studies. The initial model is a Tobit. This is censored from above and below, because of the nature of the dependent variables. Tobits are strongly affected by violations of well-behaved errors. For example, heteroskedasticity causes bias in Tobit models, but not in OLS models. So all three efficiency models are tested for heteroskedasticity. The null in each case is clearly rejected (p-values are equal to 0 in each case, where the heteroskedasticity is modelled as being caused by the same variables as in the model). The three models are then re-estimated as CLAD (Censored Absolute Deviations) models. To attempt to account for possible endogeneity fitted off-farm hours replace off-farm hours, where the fitted values come from a regression of off-farm hours on the number of main males and females in the home, the number of non-main males and females in the home and the number of family members living away from home. Standard errors are bootstrapped in two steps, over segmento and then over individuals.

Table 6-2: CLAD efficiency regressions

		Technical efficiency					Allocative efficiency					Scale efficiency					
	Reps	Observed	Bias	Std.Err.	[90% Conf. Interval]		Observed	Bias	Std.Err.	[90% Conf. Interval]		Observed	Bias	Std.Err.	[90% Conf. Interval]		
Spanish speaking	1000	-0.037	0.034	0.120	-0.234	0.161	0.021	0.004	0.080	-0.111	0.152	0.087	0.055	0.109	-0.092	0.266	N
					-0.167	0.237				-0.112	0.148				0.022	0.349	P
					-0.202	0.179				-0.124	0.141				-0.048	0.195	BC
Inverted Simpson Index	1000	0.101	-0.011	0.049	0.020	0.182	-0.063	0.003	0.025	-0.104	-0.022	-0.034	-0.022	0.049	-0.115	0.047	N
					0.006	0.172				-0.097	-0.014				-0.152	0.005	P
					0.029	0.195				-0.101	-0.021				-0.111	0.020	BC
Proportion of farm output sold on the market	1000	0.810	-0.064	0.171	0.527	1.092	0.160	0.008	0.108	-0.017	0.337	-0.071	-0.030	0.132	-0.288	0.147	N
					0.452	1.007				0.010	0.359				-0.337	0.061	P
					0.546	1.102				0.012	0.362				-0.314	0.067	BC
Years of education	1000	0.010	-0.007	0.011	-0.008	0.028	0.000	0.002	0.009	-0.015	0.014	-0.005	-0.002	0.012	-0.025	0.015	N
					-0.017	0.019				-0.012	0.018				-0.030	0.007	P
					-0.002	0.035				-0.013	0.016				-0.039	0.004	BC
Gender (1=male)	1000	-0.267	-0.030	0.149	-0.513	-0.021	0.013	0.051	0.124	-0.191	0.217	0.044	0.075	0.200	-0.286	0.374	N
					-0.537	-0.058				-0.138	0.279				-0.105	0.473	P
					-0.461	0.009				-0.231	0.168				-0.333	0.236	BC
Farm size	1000	-0.003	0.005	0.023	-0.041	0.034	-0.038	0.002	0.012	-0.057	-0.019	-0.037	-0.012	0.022	-0.074	0.000	N
					-0.029	0.048				-0.053	-0.017				-0.094	-0.022	P
					-0.028	0.048				-0.057	-0.020				-0.074	-0.005	BC
Off-farm hours	1000	-0.00038	-.00001	0.0004	-0.0001	0.00003	0.002	-.0001	0.0003	-0.0002	0.0006	0.00001	-0.000	0.0004	-0.0004	0.0005	N
					-0.0001	0.00003				-0.0001	0.0007				-0.0004	0.0005	P
					-0.0001	0.00003				-0.0001	0.0007				-0.0004	0.0005	BC
Intercept	1000	0.147	0.097	0.185	-0.156	0.451	0.509	-0.048	0.124	0.305	0.713	1.063	0.041	0.180	0.766	1.360	N
					-0.012	0.565				0.249	0.659				0.906	1.444	P
					-0.093	0.397				0.356	0.804				0.898	1.427	BC
Initial sample size		378					378					378					
Final sample size		358					378					326					
Pseudo R2		0.233					0.115					0.086					

N = normal, P = percentile, BC = bias-corrected

The initial results in Table 6-2[33] confirm the suggestions of the previous section. Crop diversification and the proportion of crops sold are associated with increased technical efficiency. They are also associated with allocative efficiency, but the diversification index has a negative sign. The only variable associated with scale efficiency is farm size. Gender is insignificant, as is education, the effect of which – from the visual information of the last section – only seems to come through at very high education levels.

One potentially important result comes in the allocative efficiency regression. 'Off-farm hours' is extremely close to being significantly positive – with a p-value of about .11 (this result is replicated when earnings substitute for hours in Table 6-A-1). This, combined with the evidence of Figure 6.6, suggests (tentatively) that one of the benefits of off-farm work (aside from directly producing cash) is on its effects on the choice of crops to produce. It appears probable that the greater security from having a separate source of cash away from the farm does have some (small) effect on allocational decisions among outputs on the farm. This set of results also strengthens the evidence from Chapter 4 concerning the role of non-farm work in insurance against risk. Finally, while off-farm work is also associated with lower technical efficiency (in Figure 6.6) the CLAD results suggest that the effects here are not significant, so the net overall effect of off-farm work on the efficiency of farm output is more likely to be positive than negative.

The second set of results in Table 6-3 are from regressions where the proportion of crops sold variable is omitted. The degree of commercial emdeddedness of farms may depend on education, gender or ethic group, among other factors. Since this variable is the most powerful variable in explaining technical and allocative efficiency it was of interest to see which other variables come to the fore if it is omitted. What happens (apart from a sharp decline in model fit) is that the ethnic group variable becomes significant. It appears that the lack of measured efficiency of non-Spanish speaking households is bound up with their lack of involvement in output markets.

[33] Table 6-A-1 in the appendix uses off-farm earnings instead of hours, to give substantially the same results.

Table 6-3: CLAD efficiency regressions without the market embeddedness variable

		Technical efficiency					Allocative efficiency					Scale efficiency					
	Reps	Observed	Bias	Std.Err.	[90% Conf.	Interval]	Observed	Bias	Std.Err.	[90% Conf.	Interval]	Observed	Bias	Std.Err.	[90% Conf.	Interval]	
Spanish speaking	1000	0.50	-0.020	0.150	0.247	0.746	0.11	0.01	0.05	0.021	0.192	0.062	0.062	0.081	-0.071	0.195	N
					0.221	0.714				0.031	0.201				0.005	0.259	P
					0.283	0.83				0.017	0.181				-0.012	0.229	BC
Inverted Simpson Index	1000	0.09	0.006	0.06	-0.016	0.202	-0.06	0.003	0.022	-0.090	-0.023	-0.043	-0.008	0.043	-0.115	0.028	N
					0.005	0.212				-0.091	-0.017				-0.137	0.001	P
					0.009	0.22				-0.099	-0.026				-0.146	-0.001	BC
Years of education	1000	0.004	-0.002	0.014	-0.018	0.027	0.002	0.002	0.008	-0.014	0.013	-0.002	-0.004	0.012	-0.025	0.018	N
					-0.019	0.026				-0.011	0.016				-0.031	0.007	P
					-0.019	0.027				-0.014	0.012				-0.008	0.018	BC
Gender (1=male)	1000	-0.295	-0.020	0.20	-0.63	.04	-0.03	-0.03	0.118	-0.073	0.32	0.075	0.039	0.183	-0.226	0.377	N
					-0.65	.014				-0.098	0.270				-0.106	0.426	P
					-0.63	0.04				-0.044	0.332				-0.149	0.359	BC
Farm size	1000	0.03	-.02	0.03	-0.017	0.085	0.004	0.004	0.011	-0.055	-0.019	-0.038	-0.011	0.023	-0.075	-0.001	N
					-0.029	0.07				-0.048	-0.012				-0.091	-0.025	P
					-0.011	0.08				-0.056	-0.023				-0.079	-0.008	BC
Off-farm hours	1000	-0.0008	0.0001	0.0004	-0.0002	0.0001	0.0001	0.0003	0.0003	-0.0003	0.0005	-0.0001	0.0001	0.0003	-0.0007	0.0005	N
					-0.0002	0.0001				-0.0002	0.0006				-0.0005	0.0006	P
					-0.0002	0.0002				-0.0003	0.0005				-0.0008	0.0002	BC
Intercept	1000	0.30	0.03	0.22	-0.035	0.69	0.5	-0.003	0.121	0.299	0.698	1.040	0.026	0.167	0.772	1.32	N
					0.012	0.73				0.264	0.661				0.905	1.36	P
					-0.028	0.68				0.322	0.747				0.932	1.47	BC
Initial sample size		378					378					378					
Final sample size		353					378					316					
Pseudo R^2		0.14					0.12238319					0.085712					

N = normal, P = percentile, BC = bias-corrected

Finally, with regard to scale efficiencies, farms are divided into those that are, in a sense, 'too big' (they have diminishing returns to scale) and 'too small' (they have increasing returns to scale). Two pairs of probits are run: one for those that are too big (with and without the instruments for off-farm work), and one for those that are too small.

Table 6-4: Regressions for scale efficiency

	Probit		IV Probit	
	'toobig'=1	'toosmall'=1	'toobig'=1	'toosmall'=1
Total off-farm hours	0.000 [0.24]	0.000 [0.92]	0.000* [0.09]	0.000 [0.51]
Inverted Simpson Index	0.101 [0.16]	-0.011 [0.92]	0.124* [0.10]	-0.033 [0.77]
Proportion of farm output sold on the market	0.698*** [0.01]	-0.244 [0.45]	0.771*** [0.00]	-0.285 [0.41]
Gender (1=male)	-0.173 [0.60]	0.646 [0.16]	-0.151 [0.66]	0.512 [0.32]
Farm size	0.119*** [0.00]	-0.609*** [0.00]	0.113*** [0.00]	-0.587*** [0.00]
Years of education	0.040* [0.07]	-0.014 [0.64]	0.035 [0.12]	-0.011 [0.71]
Spanish speaking	-0.614*** [0.00]	0.367 [0.14]	-0.687*** [0.00]	0.420 [0.12]
Intercept	-0.530* [0.07]	-0.822** [0.03]	-0.702** [0.03]	-0.57 [0.23]
N	368	368		
LR $chi^2(7)$	33.46	39.25		
Log Likelihood	-235.944	-113.074		
Pseudo R^2	0.066	0.148		

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets.

These are not obviously very informative – larger farms (and farms run by non-Spanish speaking households, which tend to have more livestock than equivalent farms run by Spanish speakers) tend to be 'too large', and smaller farms tend to be 'too small'. When we look at the raw statistics for the sierra, farms that are 'too big' are on average 1.5 hectares, compared to 1 for those that are not. Farms that are too small have a median of .2, compared to 1.2 for those that are not. This suggests a median of around 1-1.2 hectares as being the efficient size, and this is indeed what is seen in Figure 5, where scale efficiency peaks at 1 hectare.

From Figure 6-5, allocative efficiency also peaks at 1 hectare and, as with scale efficiency, declines thereafter, while technical efficiency peaks at 2 hectares, and is relatively stable thereafter. Even though neither scale nor allocative efficiency is likely to be a true efficiency measure but rather a measure of the extent to which farmers are forced to react to overcome market imperfections, it is unlikely that the declining allocative and scale efficiencies for larger farms (as opposed to smaller ones) are completely unrelated to efficiency. For a best size for overall economic efficiency, therefore, Figure 6-5 would suggest a larger size – probably around 2 hectares, where technical efficiency reaches a plateau and scale and allocative efficiency have already started to noticeably decline (farm size is not statistically significant for technical and allocative efficiency, so this estimate is a judgement call rather than a straightforward reading of the results).

6.5 Off-farm work and permanent income: introduction

Figure 6-6 of this chapter suggests that off-farm work may play a role in improved allocative efficiency. This effect did not come through significantly in the allocative efficiency regression, but the effect appears highly non-linear in the figure, and the regression signs were as expected and were very close to significant. If off-farm hours indeed improve allocative efficiency on the farm, then it is possible that this is the result of an insurance effect of having income from multiple sources, as well as through the simple effect of having extra income. The multiple source insurance effect, if it exists, should allow an increase in total expenditure for the farm family, beyond that from simply having more money now. Moreover, if there are gains in input efficiency due to off-farm work not measured in the last section, which only focused on outputs, then permanent income (as measured by total expenditure) may be further increased. In the last (brief) section of this chapter, therefore, total household expenditure is regressed on a number of variables, including on-farm diversification and off-farm diversification. The main purpose of this section is to get some purchase on the relative importance of these two possible risk management strategies for farm households. These regressions develop the similar regressions towards the end of Chapter 2 (Tables 2-11 and 2-12), but are more ambitious than those exploratory regressions; through the use of the interactive hours variables, they attempt to model the determinants of permanent income.

6.6 Permanent income regressions

The regressions in this section adapt those by Deininger and Olinto (2001). Total expenditure, as a measure of permanent income, is on the left hand side. The right hand side variables are farm size, farm capital and the language group dummy. A number of other assets (self-employed capital, livestock numbers, number of rooms in the house) were also included initially, but the two farm variables dominated. The rest of the variables are multiplicative variables. They are: hours in family agricultural work by males, hours in family agricultural work by females, hours in off-farm work by males, hours in off-farm work by females, household average education multiplied by total on-farm hours, household average education multiplied by total off-farm hours, and the two main variables of interest: the crop diversification variable multiplied by on-farm hours and the income source variable (number of income sources outside the family farm) multiplied by the number of off-farm hours.

Three regressions are run, all with fixed effects at community level. In the first, all variables are assumed to be exogenous. In the second, assets are assumed to be endogenous, and farm size is instrumented by farm size in 1991. In the third, hours of work are assumed to be endogenous (because of a possibly missing household level variable affecting both working hours and permanent income) and are instrumented using the instruments of the last section, plus the number of children 0-7 and the number of children 0-15 and the number of people over 65 in the household.

Table 6-5: Determinants of permanent income

Dependent variable	Total expenditure		
	All exogenous	Farm size instrumented	Labour instrumented
Hours*males in off-farm	-0.203 [0.21]	-0.264 [0.11]	3.882** [0.01]
Hours males on-farm	-0.164** [0.04]	-0.089 [0.27]	0.661** [0.04]
Hours females off-farm	-0.229 [0.15]	-0.285* [0.08]	3.016** [0.03]
Hours females on-farm	-0.213** [0.04]	-0.199* [0.06]	1.543** [0.05]
Diversification*on-farm hours	0.069*** [0.00]	0.063** [0.01]	-0.238** [0.01]
Non-agricultural income sources*hours off-farm	0.032 [0.16]	0.041* [0.08]	-0.354*** [0.01]
Years of education*hours off-farm	0.072*** [0.00]	0.070*** [0.00]	-0.171* [0.06]
Years of education*hours on-farm	0.039*** [0.00]	0.041*** [0.00]	0.081* [0.07]
Spanish speaking	223.66 [0.53]	221.475 [0.55]	850.309 [0.12]
Farm size	67.655*** [0.00]	110.144*** [0.00]	66.304*** [0.01]
Farm capital	0.575*** [0.00]		0.609*** [0.00]
Intercept	2,888.163*** [0.00]	3,041.559*** [0.00]	-707.225 [0.55]
N	980	980	980
Number of SEG	109	109	109
R^2	0.202		
Instruments		Farm size (1991)	No. of main persons (male), no. of main persons (female), no. of non-main males, no. of non-main females, no. of over 65s, no. of members not at home, under 15s

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets.

The third regression appears quite unreliable (education multiplied by off-farm hours reduces permanent income).The second regression is the only one that gives firm evidence of a positive effect of income source diversification on outcomes. From the first two regressions, and even in the third, it appears that on-farm diversification is important in improving permanent income. This is probably through a combination of effects – including both risk reduction and the improved technical efficiency noted in the last section from having plots at a variety of altitudes. This is a strong, and potentially quite important, result.

The results from the education variables are interesting in themselves. Increased average household education multiplied by the amount of farm work does appear to improve permanent income, but by less than the same education level with the same amount of off-farm work. The weak relationship between education and farm efficiency suggested by Figure 6-4 and close to coming through in the efficiency regression of Table 6-2 does finally appear to come through for on-farm work in this regression. The relationship appears a lot stronger with off-farm work returns, as the literature suggests. Nevertheless, the on-farm result is significant.

Gender differences, on the other hand, appear extremely small, perhaps slightly weighted towards males, but nowhere approaching statistical significance.

A small somewhat arbitrary but fairly plausible simulation using these results shows the relative effects of off-farm work (plus education) and on-farm work (plus farm assets) as potential routes out of poverty. Using the coefficients from the first model, we find that a household of males with an average of 5 extra years of education, working an extra 2000 hours off the farm, with 2 off-farm income sources, improves permanent household income by an average of 380 soles. On the other hand, a household of males with no extra education but with an extra 3 hectares of land and 750 soles worth of farm capital, with 2 crops providing equal revenue, working the same 2000 extra hours on the farm, improves household income by an average of 734 soles[34]. Allowing for the arbitrariness of this example, it is not implausible to suggest that having physical assets probably dominate human capital in their importance for income gains, unless the extra human capital is large relative to local standards.[35] The regressions of Chapter 2 (in Tables 2-11 and 2-12) suggest that the asset effects are probably larger for people with lower permanent income and lower human capital, and the human capital effects stronger for the better off generally, but also for those with very few land assets. The results of the last chapter are also relevant here: in

[34] If the extra land is only 1 hectare and the extra capital 375 soles, then the gains from the two strategies are roughly equal. It appears, therefore, that for households that diversify in this way (with 2 off-farm sources of income, or two equally valuable crops) 3 extra years of education for everyone in the household is equivalent in value to getting an extra hectare of land and 375 soles of capital (for comparison purposes, in the coastal region the average amount of capital is over 1000 soles, and in the sierra it is over 800, falling below 300 in the rainforest).

[35] As can be seen in Table 2-6, physical assets are far more unequally distributed than is human capital, so the apparently large absolute differences in farm assets used in the example are actually smaller differences (than the three year education difference of the example) if land and capital assets and education are standardised.

Table 5-6, it is reported that wage workers with an average 8.2 years of education were more likely to be unconstrained in off-farm hours than people with less than 5.3 years education. There may be threshold effects from having extra education (at some point becoming able to find unrationed work) that the linearization of the econometric model smoothes out into relatively low year on year returns. Quadratic terms were included in these regressions, but to no avail.

6.7 Conclusion

The aim of this chapter has been to measure the effects of off-farm work on agricultural efficiency and permanent income, as measured using three farm efficiency measures and household total expenditure.

With regard to farm efficiency first, Figure 6-4 suggests that there is a weak positive relationship between allocative efficiency and off-farm hours. Doing between zero and 500 or so hours off-farm work is associated with lower allocative efficiency. Between 500 and about 1500 hours there appears to be little relationship, but above about 1500 hours per annum off-farm work appears to have a positive relationship with allocative efficiency. Although this does not show through as being significant in the econometric model, it is almost significant (the 10% confidence interval is between about -.00001 and .00007). This suggests that once the rewards from off-farm work become a significant part of the family budget then gains to agricultural output are possible. With regard to technical or scale efficiency little relationship seems to exist with off-farm work.

The hiring in results in Table 5-2 of the last chapter suggest that larger farms will become less efficient because of non-linear transaction costs for hiring in. Figure 6-5 in this chapter suggests, specifically, that as farms get larger than around 2 hectares or so, in the sierra, they become less efficient. This is an important result in the land distribution debate studies (Berry and Cline, 1979; Benjamin, 1995; Barrett, 1996, Lamb, 2003). The median farm size in the sierra is 1 hectare in 1994, suggesting that some farms are slightly too small to reap maximum productivity gains (although most individual farms are seen as too big in the scale estimations).

Regarding permanent income (which may be improved, in one way, through the allocative efficiency channel), off-farm work appears to be more rewarding than on-farm work on a per hourly basis (beyond a certain amount embodied in the positive farm size and capital coefficients). This finding echoes that of the last chapter, which suggested that more than 40% of wage working individuals on mainly small farms were rationed in the amount of off-farm work available to them and would, if they could, work longer off-farm. All else equal, the number of income sources themselves available to a household only weakly increases permanent income, which suggests that the main boon of off-farm work is simply from the extra money gained (which would be more were hours not often rationed), rather than the potential insurance effect from having extra income sources (if this were so, multiple sources would probably have a stronger effect).

Considering efficiency and permanent income together, a number of conclusions emerge. Crop diversification appears to increase technical efficiency while, not unexpectedly, reducing allocative efficiency. The overall effect on permanent income appears to be quite strongly positive. An extra hour spent on 2 half-hectare plots, it would seem, are more rewarding than an extra hour spent on a single 1 hectare plot. This may be because of rapidly diminishing marginal returns to labour inputs on individual plots, even relatively large ones. It may be that adding livestock output would negate the positive effect of crop diversification on efficiency if farms with livestock tend to have fewer plots. But even if this were the case the likelihood remains that some form of agricultural diversification improves permanent income. The effect is greater within a plausible range than the insurance effect of off-farm work, as measured by the number of income sources variable. Until fairly high levels of education are reached it is also better, on average, than the direct financial effect of off-farm work (when one compares the size of coefficients, it is found that the effect of an extra hour's on-farm work of an all-male household with, say, three equally valued crops is to increase permanent income by more than the effect of an extra hour's off-farm work, for all education values less than 7 years, at which point finally the education premium for off-farm work outstrips the diversification premium for on-farm work).

Language spoken does not appear to have any direct effect on permanent income or efficiency, but related effects do appear to operate in a variety of ways. Non-Spanish speakers are less embedded in the market for agricultural outputs and because of this, it would appear, or at least related to it in some way, are less efficient at agriculture than their Spanish speaking counterparts. Non-Spanish speakers also tend to have a lower level of assets than Spanish-speaking households[36], at least with regard to farm size (they have slightly more farm capital on average, much of which comes from having livestock that can be used for traction and/or for insurance). Given the variety of market imperfections that probably exist beyond those already mentioned, the relative lack of land is almost certainly a further impediment for non-Spanish speakers to improved permanent income. A further impediment comes through the on-farm/off-farm nexus. In Chapter 4 it was noted that non-Spanish speaking households are less educated, and so less likely to participate in off-farm wage work; but even controlling for education, as the regressions of Chapter 4 showed, the language effect remains. Chapter 5 showed that non-Spanish speakers in off-farm wage work are more likely to be constrained than Spanish speakers; it is plausible that similar mechanisms (lack of training, informational and bargaining weaknesses, economically inefficient but socially efficient practices of informal labour exchange) operate to reduce participation. The customary diversification in the sierra through having plots at different altitudes appears to be one area where indigenous language speakers gain on otherwise equivalent Spanish speaking households.

[36] Of 980 households in these regressions, 576 are Spanish speaking and 404 non-Spanish speaking. Mean household expenditure for the Spanish speaking is 5254 soles, compared to 3984 for the non-Spanish speaking (medians: 4331 and 3271). Average farm size for the two groups is 4 9 and 3.25 hectares respectively (medians: 2 and 1.49). Average farm capital for the two groups is: 730 soles and 797 respectively (medians: 178 and 537). Average self-employed capital, on the other hand, is 1112 soles compared to 157 soles. Comparisons can also be made for family size (means of 5.29 compared to 5.56), crop diversification index (1.78 compared to 2.17), proportion of crops sold (.66 compared to .28, falling to .19 for the sierra), average household education (7.23 years compared to 6.46) and hours of off-farm work (1481 compared to 1223).

7

Drawing the Conclusions Together

Introduction

In this chapter the conclusions of the previous chapters are drawn together. There are three sections: first, substantive conclusions; then, methodological conclusions. In the third section possibilities for further research are enumerated and policy implications are adverted to where appropriate.

7.1 Substantive Conclusions

The substantive conclusions from Chapters 2 through to 6 are discussed under the following headings: a) the black box of the household; b) the issue of farm size and efficiency; c) the effects of off-farm work on welfare; d) the use of on-farm diversification and off-farm work for risk management; e) market embeddedness; f) education; g) ethnic differences, h) gender, i) off-farm and non-farm work, j) separability in labour markets.

7.1.1 The black box of the household

Until recently, the dominant model for framing farm household allocation decisions has been the Unitary Utility model. In this model, tastes are either ultimately determined by a dominant member, benevolent or otherwise, or a household-level social welfare function, arrived at given particular rules for aggregating preferences. The household bargaining models of the early 1980s (Manser and Brown, 1980; McElroy and Horney, 1981) have been instrumental in starting new lines of research, one of the most important of which has been the theoretical development and empirical testing of the Collective model.

The Collective model has provided a coherent theoretical framework congruent with the intuition (and, in many cases, the evidence) that the sources of household income matter and, more generally, that power relations exist within the household and that the balance of household power is affected by a variety of factors, including exogenous factors in the wider society. These can include such phenomena as inheritance customs, divorce legislation, the particularity and type of income sources, and other factors which could affect a partner's standing should the household break up. They may also include phenomena which affect utility should the partners decide not to co-operate within the household, without going so far as to break up, such as access to consumer durables or extended family or, even, the existence of a spare bedroom. Some of these phenomena are amenable to policy, and some are probably not.

The implications of using the wrong model are not generally obvious, but a very good illustration comes from Chiuri (1999). Using an Italian dataset and a model that nests both Unitary and Collective frameworks, she shows that if a Unitary model is chosen then husband and wife's labour are found to be substitutes, and the wife's wage elasticity of labour supply is positive. If a collective model is chosen, then the two types of labour turn out to be weak complements and the wife's wage elasticity of labour supply is negative. The policy implications are very different for the two sets of results. Testing of the restrictions of the nested model rejects the Unitary but not the Collective model.

This research stream is in its initial stages, and the contribution of Chapter 3 of this paper is twofold. Firstly, for the first time the Slutsky matrix approach has been used to test if the general form of the Collective model holds for families with two parents and an adult child, a not uncommon situation in farm households. The finding is that the Unitary model is generally rejected, for both two adult and three adult households, and the appropriate Collective model is not rejected using one estimator, and is rejected at 5% (but not at 1%) using another. The second main finding is that the Collective model is far less likely to be appropriate for Spanish speakers than for non-Spanish speakers. This is a finding that needs further investigating (with larger samples, and different estimators), but it could have implications for policy.

The individual level off-farm participation (and hours) models of Chapter 4 also open up the 'black box' of the household, again in ways that are consonant with sociological observations. Farm operators work mainly in the agricultural sector, for fairly low wages. Young males and females, if they work off-farm for wages, tend to work in fairly well paid areas, such as administration or teaching. But a lot of young males also work for wages on farms, and most young females, if they work off-farm, work in self-employment, usually with their mothers, often in a craft business, or selling farm products.

From these models it is possible to make certain useful deductions. Crop diversification is, as just noted, within the male realm, handled mainly by the operator and, to a lesser extent, younger males. Household cash-flow, on the other hand, tends to be handled by the farm couple.

Finally, the separability tests of Chapter 5 also contribute to opening up the black box of the household. In the last part of the chapter, these tests are carried out at the level of the individual. Within the household, male operators are more likely to be constrained in hours worked than others (though others, especially females, may well be constrained from participating in wage work in the first place), and non-main persons (mainly adult children of operators and spouses) are less likely to be constrained in hours worked than main persons.

7.1.2 Evidence for the most efficient farm size

There has been a long debate in the literature about what constitutes the most efficient farm size when markets are imperfect. If capital imperfections dominate, then the most efficient farm size tends to be large, since smaller farms will tend not to have access to capital, and thus have lower productivity. If imperfections in hired labour dominate (e.g. monitoring costs get progressively higher), then the Chayanovian 'family farm' tends to be most efficient, since larger farms will require hired labour, which is less efficient than family labour. Evidence for both kinds of imperfection have been found in the literature, and this is referenced in Chapter 5. The separability tests of Chapter 5 and the efficiency estimations of Chapter 6 give consistent answers to this question in the case of Peru.

The evidence of Chapter 5 relates to the rejection of separability of hired in labour. Rationing of such labour is also rejected, meaning that the most likely reason for non-separability is that transaction costs are non-linear. Hiring in tends to take place on larger farms with a high female/male ratio in the household, so the most probable scenario is that monitoring of hired in labour does indeed produce a cost on these farms such that separability breaks down (Murrugarra (1998) in quite a different context, finds also that separability breakdowns are more likely in rural Peru for women than men). If this is the case, and if this is the dominant market imperfection, then the Chayanovian farm is likely to be more efficient than larger farms.

The evidence of Chapter 6 partially supports this. In Chapter 6 technical, allocative and scale efficiencies are calculated for farms in the sierra, using Data Envelopment Analysis. From examining the efficiency-farm size relationship, it is clear that the most technically efficient farm size in the sierra is around 2 hectares (while the median farm size in the sample is 1 hectare). Technical efficiency stays steady at this maximum for larger farms, but other forms of efficiency decline, so 2 hectares can be taken, roughly, as an overall (economic) efficiency high point. This farm size is below the average at which labour tends to be hired in (mean size for farms that hire in is 2.3 hectares in the sierra, compared to 1.6 hectares for those that do not), suggesting that a farm size compatible with family farming is indeed probably the most efficient.

7.1.3 Effects of off-farm work on permanent income

A series of Figures and some simple regressions in Chapter 2 suggest that there are two routes out of poverty – having assets (in this case, land) and having education, which, as seen in Chapter 4, can lead to the improved opportunity of getting non-agricultural off-farm work. The pictorial evidence from Chapter 2 suggests the amount of land that a household has is a stronger predictor of per capita permanent income (measured by per capita expenditure) than are education and off-farm participation. The regressions in the same chapter suggest that the role of assets is probably stronger in improving income for the poorest group of people, and that of education is stronger for the better off (through both on and off-farm work).

There is clear econometric evidence in Chapter 6 that for many people off-farm work itself is generally welfare improving. But is it as welfare improving as having physical assets? When a comparison is made, the suggestion is that an extra hectare of land and an approximate 50% increase in mean farm capital (375 soles) for those working on the farm is as rewarding, all else equal, as an extra 3 years of education and off-farm work.

The efficiency regressions of Chapter 6 provide suggestive evidence that working off the farm for longer improves allocative efficiency in crop choice on the farm. The effect of this gain may be reduced or eliminated by the loss in technical efficiency associated with working off the farm, but – from the efficiency Figures in Chapter 6 –

the gain in allocative efficiency in those working at least 1,500 or so hours per annum off the farm appears to outweigh any other efficiency losses. This would seem to be the second channel through which having off-farm work can improve welfare (apart from providing higher returns for many people than agriculture).

7.1.4 Crop diversification

Crop diversification and off-farm work are both potential insurance strategies against price or yield collapse for certain crops.[37] In the participation and hours regressions of Chapter 4, the strongest negative relationship between diversification and off-farm work is for operators going into wage work, or larger scale self-employment. Other things being equal, if there are fewer crops, in the sense measured by the diversification index, operators are more likely to participate, and for those that already do participate, to work longer off-farm hours. Given that in many farms, at least in the sierra (which is where most of the sample comes from), crop diversification is synonymous with farm fragmentation by altitude, this means diversification is more likely than not to be the exogenous variable in this relationship. There may be some endogenity in the other regions, but in most regressions the sierra contributes more than half of all observations.

In general, as far as intra-household allocation of tasks is concerned, it seems that if operators are able to diversify, they do. This has little or nothing to do with the rest of the family, except possibly the younger male.

In Chapter 6, we see that crop diversification (farm fragmentation) appears to lead to some fairly large scale improvements in technical efficiency in the sierra. For farms where all crops are equally valuable, having an extra crop leads to an increase in technical efficiency such that farms are around 10 percentage points closer to the frontier. The loss in allocative efficiency can be read as the price paid for managing risk by diversifying crops.

---

[37] Diversification could also be motivated by the weakness of food markets in certain areas, and a negative correlation in regressions with off-farm participation could then be due to the possible positive correlation between development of markets and off-farm participation but the fact that the relationship holds true even in the random effect models in Tables 2.4 and 2.6 (albeit with a slightly smaller coefficient) suggest that efficiency gains and risk insurance are probably the main motivators to diversify. Also, there is a small positive correlation between market embeddedness and diversification, which would not be likely if the main motivation to diversify was to provide one's own food. Finally, there is no correlation whatsoever between the number of different foodstuffs eaten by a household and the degree of diversification, controlling for segmento.

Some small simulations have been made to try to quantify the gains from diversification. In Chapter 6 it is noted that up until an education level of seven years or more, an hour on-farm with 3 equally valuable crops is more valuable than an hour off-farm. Only for the better educated is diversification not so valuable.

7.1.5 Market embeddedness (proportion of crops sold)

Market embeddedness has been proxied by the proportion of farm produced crops that are sold. In Chapter 2, Figure 2-7, the association between market embeddedness and higher per capita income is very striking. Households who sell most of their crops on the market have more money than other households.

In Chapter 3, tests for the Unitary versus the Collective model suggest that for Spanish speakers in the sierra there is no difference in what the appropriate utility model is between those households who sell most of their crops on the market and those who do not. Being involved in marketing crops such as potatoes or barley (which is generally done by men) has not changed the fundamental way couples relate. Though the balance of power may change, that power per se matters does not.

In the regressions of Chapter 4, the main results relating to this variable are that a) off-farm work is less likely, the more embedded the farm is in the market and b) this relationship only holds true for all adult members of the family. Off-farm participation thus provides cash flow for houses that do not receive cash for their goods on the market, and it is everyone in the household who appears to be in charge of managing this. The instrumental variable models of Chapter 4 suggest some endogeneity in this regard – off-farm work reducing the need to go to market – but the very strong and clear relationship between off-farm work and efficiency, shown in Chapter 6, as well as the simple pictures of Chapter 2, make one doubt this; access to the discipline of markets appears to improve efficiency, raise per capita income, and to be associated with a lower demand for off-farm work, which is, as seen above, more a recourse of the poor (compared to having farm assets) and a second best method of diversifying risk (compared to being able to diversify crops). It seems very unlikely, therefore, that market embeddedness is the result of not having off-farm work; rather, the other way round.

7.1.6 Education

From the bivariate Figures in Chapter 2, it is clear that more education is associated with higher income per capita, but not as strongly associated as is larger farm size. More education is associated with longer off-farm hours, but itself not with farm size, at least not strongly. All these bivariate relations appear non-linear.

From the regressions of Chapter 4, it seems that higher education is generally likely to lead to higher participation off-farm, most likely in the non-farm wage labour category, though also in the self-employed category (for operators and spouses). More education is also generally associated with less hours worked off-farm in agricultural wage labour and in larger scale self-employment (Table 4-9), although there is no statistical effect on hours worked generally.

One of the supplementary regressions, reported in Chapter 5, section 4, shows that education matters for off-farm participation only when farms are also using hired labour. These are generally larger, better capitalised farms. The implication is that education is important for participation in better paid work (at least better than the agricultural wage), and not so for less well-paid off-farm work.

From Chapter 6, higher education improves permanent income, both in agriculture and outside agriculture. It appears that the gain in returns is higher outside agriculture. Working from the coefficients of the first permanent income regression, an extra 5 years education, working off-farm 2500 hours a year, gains 750 soles a year off the farm household. The gain for 2500 hours of on-farm labour with 5 extra years of education is 512.5 soles.

From the efficiency figures in Chapter 6 it appears that the efficiency gain within farming is highly non-linear (and perhaps for this reason does not show up in the CLAD regressions of the same chapter, though it may also be because of the confounding effects of other variables). The main benefits to farm efficiency accrue to a few very highly educated farm households (where the average education is over 8 or 10 years or so). Outside this highly educated subset, extra years of education appear to have very mixed effects on farm efficiency, which conforms with the literature.

7.1.7 Ethnic differences (as indicated by main language of household)

The only ethnic identifier in the data is the language spoken by the head of the household. The dummy used in the regressions (1=Spanish speaking and 0=non-Spanish speaking) therefore refers to linguistic ability or choice, and not to ethnic differences per se. In fact, many Amerindians do speak Spanish as their main language.

Even so, it is still worthwhile to focus on language group differences. Mean household expenditure for the Spanish speaking is 5254 soles, compared to 3984 for the non-Spanish speaking. Mean farm size for the two groups is 4.9 and 3.25 hectares respectively (medians: 2 and 1.49). Education levels and hours in off-farm work are also lower for the non-Spanish speaking, while family sizes are larger.

The Collective model test of Chapter 3 suggests that the Collective model is more likely to apply to non-Spanish speaking families than to Spanish speaking families. From Chapter 4, it is clear that much of the farm work that is declared as unpaid is carried out by indigenous language speakers. From Chapter 4 it is also clear that speaking Spanish increases the chances of working off the farm.

The separability tests in Chapter 5 suggest (as in Vakis *et al.* 2004) that not speaking Spanish is likely to result in being rationed if one works off-farm for wages. From Chapter 6 it is clear that people who do not speak Spanish at home are also far less allocatively and technically efficient than those who do, as is seen in the Figure 6-3. There is a high correlation between language group affiliation and market embeddedness (being Spanish speaking more than doubles the proportion of crops sold). For some indigenous people, therefore, changing language and becoming more embedded in the market may be part of the same acculturation process. In the efficiency regressions, unlike in Figure 5-3, both variables are controlled for, and the market embeddedness variable comes through; only when this is dropped is the language dummy significant.

Unless the degree of market embeddedness is simply a choice made by indigenous groups, it is likely that these results can be explained by barriers to markets being in place, making it more difficult for indigenous groups to participate. This issue has been explored by Vakis *et al.* (2002) where they argue that transaction costs, in the form of infrastructural weaknesses and, possibly more importantly, informational asymmetries (about price ranges and fluctuations in different local and semi-local markets) reduce market participation (as well as the gains from markets already participated in). The language dummy is significant in a number of their regressions, and one of their policy recommendations to reduce transaction costs is improved language skills (meaning greater bilingualism – not giving up their own language).

However, whether participation in markets is enough to improve efficiency on the farm is not clear. Traditional labour exchange practices appear to reduce the chances of indigenous language speaking households hiring in farm labour (Chapter 5). These practices have been linked with reductions in efficiency (van der Ploeg, 1990). Changing these practices may improve efficiency (at some cost, perhaps, in lost pleasure or social capital); and it is quite likely that exposure to markets will affect these practices anyway. It may be, though, that the causes of the negative results for efficiency for both language group and market embeddedness variables are part of a particular over-determination of social reality, where marginalisation, remoteness, language, market embeddedness, land quality, farm size etc, all contribute to the differences found between language groups in efficiency and, through these, in permanent income.

7.1.8 Gender

On-farm tasks are distributed differently to men and women. From the regressions of Chapter 4 it is apparent that crop diversification is the preserve mainly of men. Jacoby (1991) says that, in the sierra, women look after the livestock mainly, either on their own or sharing this task with men. Women are more likely to be declared unpaid farm workers than are men (in particular, women in non-Spanish speaking households), and women are a good deal less likely to work off-farm than are men. Wages and returns to self employment are likely to be lower for women than for men. There is one exception, where young sole trading females get better returns, on average, than do (the very few) young sole trading males.

The same regressions of Chapter 4 suggest that women who do participate off the farm are more likely to enter into self-employed labour than wage work. This is true both for the spouse and for daughters, as is clear in Table 2-7. It is interesting that young males hardly ever go into self-employment (only 17 out of 201 off-farm working non-main males work in self employment). Males' work off the farm is mainly in wage work in farming, but also in building and construction (and in fisheries and forestry in the rainforest region). Females in self-employment work mainly in the selling, craft and service sectors. Younger better educated females who work for wages (there are not too many) are quite likely to be professionals e.g. teachers, secretaries.

The effect of un-earned income on the spouse's participation is strong (this is a common result in the literature), suggesting a high shadow value for her time in the home. Having young children reduces her chances of entering into self employment, but this is the only significant effect of the children variable in all the participation regressions.

The separability regressions in Chapter 5 suggest that hiring in is more common in households with a high female to male ratio, and – not surprisingly – in larger farms with smaller households. With regard to off-farm work, females who do work off-farm for wages appear to be less constrained than males, but there are not many who do so and this finding may not be very robust.

Finally, the efficiency and permanent income regressions in Chapter 6 find that female labour is generally as efficient as male labour, raising permanent income by the same amount, or perhaps only a little less so. From the regressions in Chapter 6, there is very close to being a positive statistical effect of having relatively more females in the house with regard to technical efficiency. Likewise, in the permanent income regressions, while nothing is conclusive, and there are definitely no statistical differences, there are suggestions that female labour is more productive than male on the farm, while male labour is slightly more productive off it.

7.1.9 Determinants of off-farm and non-farm work, and earnings shares

Among the main determinants of off-farm work are education (positive), and land and capital (negative).

Age has the familiar concave pattern; and on-farm experience is strongly and negatively associated with participation. The effect of unearned income is to reduce female participation in self-employment, and not to increase male participation. There is also a negative sign for male entry into large scale self employment (Table A2, Chapter 4). A positive coefficient on this variable for the large scale self-employment category would be likely were capital constraints very pressing. The wealth variable, number of rooms, has a positive sign, and may be partly endogenous.

Labour demand is an important determinant of off-farm participation: areas that are non-agricultural and with good access conditions tend to lead to more participation and longer hours, suggesting that there is probably a repressed demand in more remote areas for off-farm work.

From Chapter 2, it is clear that the share of off-farm returns declines as families get richer. Off-farm hours do decline with farm size, but per capita expenditure tends to rise (albeit non-linearly), and the share of off-farm and non-farm income also rises, slightly. Thus, returns to off-farm work tend to increase with farm size.

The share of off-farm returns is over 40% for the poorest per capita expenditure quintile, and about 36% for the richest quintile. There is a weak, but significant, negative relationship also between the number of off-farm income sources and per capita expenditure. There is no pairwise relationship at all between the number of hours worked off-farm and per capita expenditure. The fact that shares fall as permanent income rises and hours stay the same suggests, as the permanent income regressions in Chapter 6 appear to confirm, that a combination of farm assets and farm labour (at the scales simulated) is generally a more direct route out of poverty than is off-farm labour.

7.1.10 Constraints on off-farm work and on hiring-in

Separability does not appear to hold in either labour allocation decision – working off-farm or hiring in. In the first case, rationing of off-farm labour appears to be the main cause of non-separability. In the second case, non-linear transaction costs appear to be the main cause. When the exploratory latent class procedure of Vakis *et al. (*2004) is applied to the data on an individual level it appears that about 43% of the sample of wage workers are rationed in off-farm labour (this falls to 24% at household level). These are mainly badly paid workers, in sectors such as skilled construction work.

The implications of the hiring-in separability result have been discussed in 7.1.2.

7.2 Methodological innovations

The main methodological innovations are (i) the use of individual level data where it tends not to be that common: in the participation regressions in Chapter 4 and in the separability tests in Chapter 5, (ii) testing the Collective Model restrictions, using the Slutsky approach, for three people and testing for Spanish speaking households versus non-Spanish speaking households (ch. 3), (ii) using a new test to distinguish between rationing and transaction costs as drivers of non-separabilties and (ii) using CLAD models to focus on the relationship between allocative efficiency (as opposed to technical efficiency) and off-farm work.

7.2.1 The use of individual level data in participation and hours regressions (ch. 4) and in separability regressions (ch. 5)

Generally, in participation and hours models, data at the level of the operator or the two individuals in the couple or at the level of the household, is used for estimations. Many of the regressions in Chapter 4 follow on from Malchow-Møller and Svarer (2002) and use a multinomial specification where the latent variable is desired labour supply, rather than utility. Interfamilial interactions are modelled by including the averages of others in the house (e.g. average education of others) and by a specification of either a random effects model, or of household level clustering of the error terms. These are very simple reduced form models, in general, but quite a lot of information on participation and hours can be deduced from them.

A multivariate probit is also used on one occasion to model the interdependencies in the intrafamilial decision making process. The results are weak in terms of significant variables, but they do suggest that the interdependencies between operator and spouse and spouse and adult children are stronger than those between adult children and the operator.

The use of individual level data in the final separability regressions of Chapter 5 is also new. Generally, reduced form tests of this type use household level data (as in the rest of the chapter). It was felt that the exploratory nature of the latent class model needed individual level data because of the possibility of confounding regimes if the data was at the household level. The results of the individual level model seem fairly successful.

Finally, it is noteworthy that every restriction, either on the equality of coefficients across family members, or in aggregating different kinds of participation options, or in pooling data across households or segmentos, was always rejected. This means that simplicity, in the form of uniformity of coefficients across members and job categories, is never – statistically – the best option.

7.2.2 Testing the collective model for three person families (and for families of different cultural background) using the Slutsky Matrix approach (ch. 3)

The Slutsky Matrix approach to testing the Collective Model (Browning and Chiappori, 1998) has only been used once in the literature. It has two advantages over other approaches – no information is needed on exclusive or assignable goods and it tests the general version of the model, where each person's consumption may have externalities that affect the others', rather than the more restrictive 'caring' version. The three person restriction is derived and tested for the first time. The estimator used for the three person step is a single step entropy estimator, where the highly non-linear restriction is fairly easy to apply. The two person restrictions are tested using a conventional two-step estimator and the entropy estimator, with similar results for the whole data sample and for the sample when restricted to two adult households. As already noted, in 7.1.1, the Unitary model is rejected and the Collective model is not (at 1%), for both two and three person families.

7.2.3 A New test to distinguish rationing from non-linear transaction costs as determinants of non-separablity (ch. 5)

The test used in Chapter 5 to distinguish rationing from non-linear transaction costs as a driver of non-separability is a simple variant of other separability tests that use the Wu-Hausman procedure (or a Chow test procedure, also used in this case) to test for exogeneity. For an individual who is working on their own farm and also off-farm for a wage, non-separabilities may be due to rationing (a quota imposed on the amount of off-farm work available at a given wage, due possible to seasonal labour, to administrative fiat, or maybe to weak demand) or to non-linear transaction costs (costs that rise with time spent participating, such as the rising cost to the participant of being away from home for longer if there are no substitutes for home or farm work, or higher travel costs at different times of the year). The essential difference for the participant is that his or her shadow wage determines the balance between on-farm and off-farm work in the second case, whereas the ration does in the first. Total labour time will be determined by the shadow wage in both cases.

Because off-farm hours are exogenous in the first case, and not in the second, a test for the exogeneity of off-farm hours should distinguish between the two phenomena, once the lack of separability is taken as given. The tests in Chapter 5 find that rationing exists in the off-farm labour market, but that non-linear transaction costs (in this case probably supervision costs) exist in hiring in labour.

7.2.4 Examining the relationship between allocative efficiency and off-farm work (ch. 6)

Allocative efficiency is rarely measured in studies of farm efficiency, the usual focus being on technical efficiency. In developing countries allocative inefficiency can often be regarded as an efficient response to missing goods, land, capital, labour or insurance markets. If off-farm work is available, then this may mitigate the need to rearrange outputs to provide insurance. Throughout the thesis the negative relationship between on-farm diversification and off-farm participation has been noted. Diversification is found to decrease allocative inefficiency, as one might expect while greater off-farm hours go together with greater allocative efficiency. Exogeneity is difficult to be sure about in this context as all the variables are inter-related (technical efficiency, diversification, allocative efficiency and off-farm hours and earnings), but off-farm hours are instrumented in the CLAD regression and come very close to significance.

7.3 Possibilities for further research and implications for policy

This thesis uses results from and is intended to contribute to ongoing research in at least 6 areas connected with rural households in non-OECD countries. These areas are (i) research into the Collective household utility model (Lundberg *et al.* 1997; Browning and Chiappori, 1998; Zhang and Fong, 2001; Beninger *et al.* 2002; Chiappori *et al.,* 2004; Donni, 2004; Vermuelen, 2005), (ii) intra-household time allocations in rural households (Newman and Gertler, 1994; Fafchamps and Quisumbing, 2000); (iii) the management of risk in a Collective framework, (Duflo and Udry, 2001; Goldstein, 2004); (iv) separability, rationing and transaction costs in labour allocation (Benjamin, 1992; Murrugara, 1998; Fafchamps and Quisumbing, 2000; de Janvry and Sadoulet, 2004; Vakis *et al.,* 2004; Henning and Henniknsgen, 2005), (v) efficiency studies in Latin American farming (Pinheiro and Bravo-Ureta, 1997; Filho and de Souza, 2003) as well as (vi) the burgeoning literature stream on household strategies reviewed in Chapter 1, whose main exponents are probably Barrett and Reardon (2000).

Within the literature on the Collective model, the tests of Chapter 3 suggest a) that it is possible to test for the general version of the model in 3-adult families and b) that culture and ethnicity may influence the choice of appropriate models in particular ways. What the research in this paper has not dealt with, and what is being dealt with in the literature at the time of writing, is how to incorporate non-decision makers (children) into the Collective model (Chiappori *et al.,* 2002), how to model non-participation (Donni, 2001), how to deal with endogeneity of the balance of power (Koolwall and Ray, 2002). As seen in the literature review chapter, sharing rules have been estimated in a variety of settings, but not in a specifically agricultural setting. It has been suggested (in the Appendix to Chapter 1) that it should be relatively easy to adapt the Collective model theory on household labour (housework) as it exists to farm production, which would actually be more likely to meet the more stringent separability assumptions needed to incorporate this kind of production into the model. Thus, estimation of shares of agricultural profits should be possible, given the appropriate data, along with shares from wages and self-employed earnings and housework. Household heterogeneity of unknown form is certainly going to be an important complicating issue in identifying shares (Blundell *et al.*, 2002) and it has barely made its way into the literature. It would certainly need to be modelled before such estimated shares could be taken too seriously.

It should be worth trying to do this because, from a policy point of view, understanding the inner workings of households can be extremely important. Haddad *et al.* (1997) give four reasons why this might be so: (i) the effect of public transfers may differ depending on who gets the transfer; (ii) the response of non-recipients needs to be considered, (iii) not only physical resources or money but information also may not be completely shared within the household (this has been especially important, they argue, in sub-Saharan Africa, in relation to project delivery and policy implementation) and (iv) being aware of intra-household issue opens up other policy levers for governments to use (apart from transfers) – such as inheritance law, divorce and alimony rights, gender specific education policies. To these four reasons should probably be added a fifth: irrespective of policy, both agricultural and self-employed productivity can change through advances in technology, or through changes in crop choice or technical efficiency brought about by exposure to the market – it can be important to be aware who the most likely gainers and losers from such changes are.

Research into time allocation in rural households is also continuing outside the framework of the Collective model. Kim and Zepeda (2004) for Wisconsin contribute an article suggesting that intra-household allocation in rural US households is gender specific. Newman and Gertler (1994) and Fafchamps and Quisumbing (2000) suggest that different members of the household are weighted differently, as far as leisure time allocation is concerned. The latter also suggest that cultural norms may mean that efficiency is not achieved, even though there are gains in specialisation from off and on-farm work. The research in this project has not focussed much on time allocation per se, but participation in different activities. Activities have been broken down in a particular way (into paid and unpaid farm work, wage farm and non-farm work, sole trader and employer self-employment and housework), but improved time allocation data is becoming available (e.g Guatemala LSMS, 2000) and the influence of institutions and norms on intra-household allocation of more specific tasks can begin to be measured more accurately. This is important from a policy point of view because it is important to understand if certain specialisms (say, gender specific tasks or roles) have efficiency implications. Chapter 4's results show that females, especially female spouses, are more likely to enter self-employment than wage work. But reported returns from self-employment, for example, are very low according to the responses in the Peruvian data, as seen in Chapter 2. The effect on permanent income of female off-farm work, the vast majority of which is in self-employment is, however, indistinguishable from that of males in the regressions of Chapter 6, casting doubt on the reported returns figures. From Chapter 4 also, it seems likely that crop diversification is a task mainly of males in the households, while managing cash-flow is the responsibility of all adult members. Measuring who does what within the household and at the household-market interface, and how efficient they actually are at it, and what would make them either more efficient at what they already do or able to choose something else more rewarding is likely to be an important area of future research, as well as being important for policy.

One of the implications of intra-household Pareto Efficiency is that risk pooling should be complete within the household. If income pooling is an implication of the Unitary model[38], then risk pooling is of the Collective model. Individual consumption within the household should not be affected by short term individual income shocks, once total expenditure has been conditioned on. This implication of the Collective model has been tested recently (Duflo and Udry, 2001; Goldstein, 2004), and rejected, and is the main avenue through which the Collective model has been applied to rural developing economies. It is not necessary that the Collective model operate for all spheres in which household decisions are made: it is quite possible that different models are appropriate say for consumption of everyday consumer goods (as tested for in this thesis), consumer durables and long-term labour supply decisions. It is also possible that different types of households operate under different models for the same sets of decisions, as is suggested for retirement decisions in Norway by Zhiyang Jia (2005) and suggested in this thesis for everyday consumption decisions made by Spanish speaking versus non-Spanish speaking families in Peru. The findings against risk pooling in the literature, therefore, are not necessarily findings against the Collective model in spheres other than the joint management of short term income shocks. However, it is an interesting area that remains to be further explored in the wider context of household strategies, where income comes from both inside and outside agriculture. The suggestion from Goldstein (2004) is that one might need to go outside the family to analyse risk management: friends and community seem to be more important than family as helpers towards consumption smoothing for individuals in Ghana who have suffered short term income shocks. Whether the Collective model holds for risk management in non-African settings has not yet been tested for. The nature of family interaction is certainly different in Latin America from West Africa, and further research in to this issue is important.

The strain of research into separability tests in labour allocation (largely in the train of Benjamin (1992); and exemplified by Jacoby (1993) and Morrugarra (1998) for Peru), and – more recently – in trying to understand and measure inter-household heterogeneity and what might have led to separability breaking down (Hobbs, 1997; Were, 1998; Key *et al.,* 2000; Vakis *et al.,* 2002; Vakis *et al.,* 2004) has been used

[38] ....as it is usually taken to be, though Browning, Chiappori and Lechine, 2004, argue that the Slutsky matrix conditions tested in Chapter 3 are in fact the signature of a Collective model; they argue that not pooling income is possible in a Unitary context.

principally in Chapter 5 of this thesis. Small contributions have been made to the first part of this literature – a simple exogeneity test – and to the second part – using individual data and a more refined sample in the procedure developed by Vakis *et al.* (2004). This non-separability literature is part of a much wider literature on market imperfections in a variety of markets to do with agriculture (among them the markets for capital, insurance, land purchase and leasing, and non-labour inputs) in middle income and developing countries. Surveys and procedures are being developed (reported on in some of the above references) to probe at underlying causes of non-separabilities, at the linking of imperfections in different markets, and to get at the costs. Costs of non-separabilities in labour allocation have been estimated in full farm household calibrated models (Taylor and Adelman, 2002) with many households, but these depend heavily on model specification and parameters in the literature that are used in the calibration. Inter-household linkages and linkages with other market failures are not modelled. More detailed econometric work is certainly needed in this area to improve parameters in policy-influencing models.

The fifth strand in the literature to which this research is related is the measurement of agricultural efficiency, especially in Latin America (Pinheiro and Bravo-Ureta, 1997; Filho and de Souza, 2003). The results here complement some of the more general results in the literature. Education appears to have little impact on efficiency, except at the very highest levels (more than 10 years or so), and does not come through as being statistically significant in the econometric estimations. Farm size is not statistically significant either, except in the scale efficiency and the probit regressions. From these, and from the visuals, there are suggestions that the most efficient farm size is slightly larger than the actual median (about 2 hectares, in the sierra, compared to a median of 1 hectare; albeit that in the sierra many individual farms are 'too big'). However, these models do not incorporate standard errors, nor do they account for the non-independent nature of the variables in estimation, although censoring is accounted for. These are not easy areas to develop, and research is ongoing in the area. The principal objective of the initial research proposal for this thesis was to look at the relationship between off-farm hours, or earnings, and allocative efficiency. There is a suggestion that higher off-farm hours increases allocative efficiency, while the alternative risk management strategy (crop diversification) reduces it, but increases technical efficiency. These are developing areas in the efficiency literature for Latin America and there would appear to be great scope to look at the different payoffs for different strategies as the literature matures.

Finally, the sixth relevant literature strand is that on household income strategies, much of which was reviewed in Chapter 1. As noted in Chapter 1, shares of rural non-agricultural income may rise or fall with farm size (falling generally in Latin America), but tend to rise with household wealth and household income; levels of rural non-agricultural income, on the other hand, rise more often with farm size, and universally with household wealth and income (Reardon *et al.* 2001; Barrett *et al.*, 2001). The evidence of this thesis, in Chapter 2, is that shares of income from off-farm employment are not strongly related to farm size in Peru, but are (slightly) negatively related to total household expenditure (which, in turn, is related quite strongly to farm size). Education seems to be the other key variable, apart from farm size and capital assets, in turning labour into income. Education increases the chances of off-farm work, and increases the returns to both on and off-farm work. From a couple of simple calculations in section 7.1.3, it appears that, for common ranges, land and farm assets are more valuable for many people than education in improving income, but both do so. Research in this area, of heterogeneous, mainly endowment determined, household income strategies is expanding enormously, and has great potential policy significance. It links in with related research in the ability of market liberalisation (as has happened in Peru since the early to mid 90s) to diminish the effects of initial farm or land asset inequalities (Carter and Zimmerman, 1998) and with the possible re-enforcement of such inequalities by off-farm outcomes (Reardon *et al.* 2001). From the cross-section studies of this paper it appears that well paid off-farm work is not closely related to farm size, and that in fact land and education are probably alternative routes out of poverty (unlike in east Africa, for example, where education increases off-farm returns, which feed back to increase farm returns through on-farm investments; and where access to education is often dependent on initial farm assets). There is some suggestion, however, in Appendix 2 of Chapter 4, that children in households that have been growing over time are more likely to be better educated and to work off-farm outside agriculture. This suggests that in the long term improving access to land may enable many landholders' descendents to leave it for better paying off-farm work. Another suggestion along the same lines emerges from Table 4-A.6, where a small group of farmers in large scale self-employment have farms that are expanding; the suggestion is that their self-employment could be funding the farm expansion. These, and a lot of other related questions, are very important issues for future research.

7.4 Conclusion

In this thesis a single dataset has been used to investigate the relationships between individual, household and farm attributes in rural Peru and a range of issues connected to decision making in general, and to income generating decisions in particular. The thesis has used the neo-classical farm household model as its organizing principle and the investigation has been carried out using quantitative techniques of visual data analysis, econometrics and linear programming. Insights into the inner working of households have been gained and the issues of constraints on choice and determinants of efficiency examined. Along the way a picture of the farm household in different regions within Peru has been filled in. A number of methodological issues have been discussed, and some gains made.

However, an enormous amount of further research needs to be done to improve understanding of relationships within households and of the links between households, communities and their wider regions. This thesis has tried to elucidate some of the issues at a very micro level and to tell some small stories. Broader developments in technology, and increasing trade and growth and their efficiency and distributive effects are ultimately what will drive change, but knowledge at the household level, and within the household, is also crucial. Differences that do not seem very important (except to those involved) may become important at a more macro scale. This could apply to phenomena as diverse as the effects of culture on how family decisions are made, the resourcefulness needed to farm at different altitudes, the skills learned in small scale craft self-employment, the social capital built up through voluntary labour exchange, the costs of monitoring hired labour or the losses accrued by not having access to education. Neo-classical economics provides a lens through which these phenomena can be viewed, one that has advantages and disadvantages and which can never provide a complete picture. But the application of its quantitative tools will often yield insights that turn out to be interesting, and often useful, either directly, or indirectly through improvement by others. To achieve some of these insights has been the ambition of this thesis.

Bibliography

Abbot, M. and O. Ashenfelter (1976). "Labour Supply, Commodity Demand and the Allocation of Time". *Review of Economic Studies*, 43: 389-411.

Abdulai, A. and A. Crole-Rees (2001). "Determinants of Income Diversification amongst Rural Households in Southern Mali". *Food Policy,* 26: 437-452.

Abdulai, A. and P. Regmi. (2000). "Estimating Labour Supply of Farm Households under Non-Separability: Empirical Evidence from Nepal". *Agricultural Economics,* 22:309-320.

Adams, R.H. and H. Alderman (1992). "Sources of Income Inequality in Rural Pakistan: A Decomposition Analysis". *Oxford Bulletin of Economics and Statistics,* 54(4):591-608.

Alasia, A. and C. Soregaroli. (2002). "Economic Growth and Food Demand Changes in Mozambique: An Assessment of Scenarios and Implications for Policy". Centre for International Development. Bologna. WP-N-3.

Alberti, A.M. (1988). "From Recommendation Domains to Intra-Household Dynamics and Back: Attempts at Bridging the Gender Gap" in *Gender Issues in Farming Systems Research and Extension.* Susan V. Poats, M. Schmink and A. Spring (eds.). Boulder Colo. Westview Press. pp 61-72.

Altonji, J., S. Hayashis and L. Kotlikoff (1989). "Is the Extended Family Altruistically Linked? Direct Tests using Micro Data". NBER Working Paper no. 3046. New York.

Arayama, Y. (1986). "Time Allocation of Japanese Farm Households". Unpublished PhD dissertation. University of Chicago.

Ashenfelter, O. and Heckman J.J. (1974). "The Estimation of Income and Substitution Effects in a Model of Family Labour Supply". *Econometrica*, 42: 73-84.

Ashworth, J. and Ulph (1981). "Household Models". In *Taxation and Labour Supply*, C. Brown (ed) London.

Barnett, W.A. (1979). "The Joint Allocation of Leisure and Goods Expenditure". *Econometrica,* 47: 539-563.

Barnum, H.N. and L. Squire (1979). "An Econometric Application of the Theory of the Farm Household". *Journal of Development Economics,* 6: 79-102.

Barrett, C.B. (1997). "How Credible are Estimates of Peasant Allocative, Scale or Scope Efficiency? A Commentary". *Journal of International Development,* 9.(2): 221-229.

Barrett, C.B. and T. Reardon (2000). "Asset, Activity and Income Diversification Among African Agriculturalists, Some Practical Issues". Cornell. Ithaca. Working Paper, 00-19.

Barrett, C.B., T. Reardon and P. Webb (2001). "Nonfarm Income Diversification and Household Livelihood Strategies in Rural Africa: Concepts, Dynamics and Policy Implications". *Food Policy,* 26: 315-331.

Bar-Shira, Z., R.E. Just, D. Zilberman (1997). "Estimation of farmer's risk attitude: an econometric approach". *Agricultural Economics,* 17:211-22.

Basu, K., Gender and Say (2004). "A Model of Household behaviour with Endogenously Determined Balance of Power". Harvard Institute of Economic Research Discussion Paper no. 2054.

Becker, C.M., E. Musabek, A. Seitenova and D. Urzhumova (2003). "Short Run Responses of Men and Women during a Period of Economic Turmoil". *Eurasian Geography and Economics*, forthcoming.

Becker, G. S. (1981). "A Theory of Social Interactions". *Journal of Political Economy,* 82:1063 – 1093

Benjamin, C. (1994). "The Growing Importance of Diversification Activities for French Farm Households". *Journal of Rural Studies,* 10:331-42.

Benjamin, D. (1992). "Household Composition and Labour Demand: A Test of Rural Labour Market Efficiency". Princeton University. Industrial Relations Section Working Paper No. 224.

Benjamin, D. (1995) Can Unobserved Land Quality Explain the Inverse Productivity Relationship? *Journal of Developmentv Economics* 46: 51-84

Berdegue, J, A., E. Ramirez, T. Reardon and G. Escobal (2001). "Rural Nonfarm Employment and Incomes in Chile". *World Development,* 29(3): 441-425.

Berry, A. and W. Cline (1979). "Agrarian Structure and Productivity in Developing Countries". Baltimore: The John Hopkins University Press.

Bhalla, S. and P. Roy (1988). "Mis-specification in Farm Productivity Analysis: The Role of Land Quality". *Oxford Economic Papers,* 40(March):55-73.

Bhalotra, S. (2000). "Is Child Work Necessary?" Suntory and Toyota Centres for Economics and Related Disciplines. Development and Economic papers 26.

Bhalotra, S. and C. Heady (2000). "Child Farm Labour: The Wealth Paradox". Bristol Discussion Paper no. 00/492.

Bhattacharyya, A. and S. Kumbhakar. (1997). "Market imperfections and output loss in the presence of expenditure constraint: A generalised shadow price approach". *American Journal of Agricultural Economics*, 79(3):860-71.

Binswager, H.P. (1980). "Attitudes toward Risk: Experimental Measurement in Rural India". *American Journal of Agricultural Economics,* 62:395-407.

Blacklow, P. and R. Ray (2003). "Intrahousehold Resource Allocation, Consumer Preference and Commodity Tax Reforms: The Australian Evidence". The Economic Record. 79 (247) 425-433.

Block, S. and P. Webb (2001). "The Dynamics of Livelihood Diversification in Post-Famine Ethiopia". *Food Policy,* 26: 333-350.

Blundell, R. and Walker I. (1986). "A Life Cycle Consistent Empirical Model of Family Labour Supply using Cross Section Data". *Review of Economic Studies,* LIII, 539-558

Blundell, R., P.A. Chiappori, T. Magnac and C. Meghir (2002). "Collective Labour Supply: Heterogeneity and Nonparticipation". The Institute for Fiscal Studies WP01/19.

Blundell, R.W. and Walker I. (1982). "Modelling the Joint Determination of Household Labour Supplies and Commodity Demands". *Economic Journal*, 92: 351-364.

Bollman, R.D. (1991). "Efficiency Aspects of Part-time Farming". in Hallberg, M. C., Findeis, J. L., and Lass, D. A. (eds.) Multiple Job-holding among Farm Families, Iowa State University Press, Ames.

Bourguignon, F. and M. Fournier and M. Gurgand (2001). "Selection Bias Correction Based on the Multinomial Logit Model". CREST Working Paper.

Bourque, S. and K.B. Warren (1981). "Women of the Andes. Patriarchy and Social Change in Two Peruvian Towns". Ann Arbor. University of Michigan Press.

Bowen, W.G. and T.A. Finnegan.(1969). "The Economics of Labour Force Participation". Princeton N.J.. Princeton University Press.

Bowlus, A. and T. Sinclair (2003). "Moving Towards Labour Markets? Labour Allocation in Rural China". *Journal of Development Economics,* 71(2):561-583.

Brick, A., E. Garvey and M. Cuddy (2005). "Off-Farm Participation in Co. Mayo". Department of Economics, National University of Ireland Galway. Working Paper Series 102.

Browning, M. and C. Meghir (1991). "The Effects of Male and Female Labour Supply on Commodity Demands". *Econometrica,* 59(4): 925-951

Browning, M. and Chiappori P.A. (1998). "Efficient Intra-Household Allocations: A General Characterization and Empirical tests". *Econometrica,*66 (6): 1241-1278.

Browning, M., A. Deaton and M. Irish. (1985). "A Profitable Approach to Labour Supply and Commodity Demands over the Life Cycle." *Econometrica,*53 (3): 503-544.

Browning, M., F. Bourguignon, P.A. Chiappori and V. Lechene (1994). "Income and Outcomes: A Structural Model of Intrahousehold Allocation". *The Journal of Political Economy,* 102(6): 1067-1096.

Brummer, B. (2001). "Estimating Confidence Intervals for Technical Efficiency: the Case of Private Farms in Slovenia". *European Review of Agricultural Economics*, 28 (3): 285-306

Cai, J. (1989). "Are Husbands and Wives Altruistically Linked? Evidence from Micro Data". mimeo Boston University.

Canagarajah, S., C. Newman and R. Bhattamishra (2001). "Non-farm Income, Gender and Inequality: evidence from Rural Ghana and Uganda". *Food Policy,* 26: 405-420.

Carter, M.R. and K. Wiebe (1990). "Access to Capital and its Impact on Agrarian Structure and Productivity in Kenya". *American Journal of Agricultural Economics,* 72: 1146-50

Carter, M. R. (1997). "Environment, Technology and the Social Articulation of Risk in West African Agriculture". *Economic Development and Cultural Change* vol. 45 (3): 557-590. University of Chicago.

Carter Michael R. and Frederic Zimmerman. (1998) "The Dynamic Cost and Persistence of Asset Inequality in an Agrarian Economy". University of Wisconsin Madison. Staff Paper no. 416.

Carter, M. R. and P. Olinto (2000). "Getting Institutions 'Right' for Whom: Credit Constraints and the Impact of Property Rights on the Quantity and Composition of Investment". University of Wisconsin-Madison. Staff Paper 433.

Carter, M. R. and Yao Y. (2002). "Local versus Global Separability in Agricultural Household Models: the Factor Price Equalization Effect of Land Transfer Rights". *American Journal of Agricultural Economics,* 69(2): 369-377

Charnes, A., W.W. Cooper and E. Rhodes (1978). "Measuring the Efficiency of Decision Making Units". *European Journal of Operations Research,* 46: 185-200.

Chavas J.P and Aliber M. (1993) An Analysis of Economic Efficiency in Agriculture: a Non-Parametric Approach. *Journal of Agricultural and Resource Economics,* 18. 1-16.

Chavas J.P. R. Petrie and M. Roth. (2005) "Farm Household Production Efficiency. Evidence from the Gambia" *American Journal of Agricultural Economics 87(1) feb. 160-179*

Chayanov A.V. (1926) The Theory of Peasant Economy ed. By D. Thorner et al. Homewood Illinois. Richard D. Irwin, 1966.

Chen, Z. and L. Kuo (2001). "A Note on the Estimation of the Multinomial Logit Model with Random Effects". *American Statistician*, 55:89-95.

Chiappori, P.A. (1988). "Rational Household Labour Supply". *Econometrica,* 56(1): 63-90.

Chiappori, P.A. (1992). "Collective Labor Supply and Welfare". *Journal of Political Economy,* 100:437 – 67

Chiappori, P.A., B. Fortin and G. Lacroix (2001). "Marriage Market, Divorce Legislation and Household Labour Supply". CIRANO Working Papers 2001s-16.

Chiappori R., Blundell R. and Meghir C. (2002) "Collective Labour Supply with Children". Institute for Fiscal Studies. Working Paper W02/08.

Chiuri, M. (1999). "Intra-Household Allocation of Time and Resources: Empirical Evidence on a Sample of Italian Households with Young Children". Centre for Studies in Economics and Finance. Working Paper no. 15.

Cloutier, L.M. and R. Rowley (1993). "Relative Technical Efficiency: Data Envelopment Analysis and Quebec's Dairy Farms". *Canadian Journal of Agricultural Economics,* 41: 169-176.

Coelli, T.J., R.D.S. Prasada and G.E. Battese (1998). "An Introduction to Efficiency and Productivity Analysis". Kluwer Academic Publishers, Boston/ Dordrecht/London.

Corral, L. and T. Reardon (2001). "Rural Nonfarm Incomes in Nicaragua". *World Development,* 29(3): 427-442.

Couprie, H. (2003). "Time Allocation within the family: Welfare Implications of Life in a Couple". GREQAM. Working paper 02A23

Da Silva, G. and M. Del Grossi (2001). "Rural Non-farm Employment in Brazil: Patterns and Evolution". *World Development,* 29(3): 443-453.

de Janvry, A. and E. Sadoulet (2001). "Income Strategies Among Rural Households in Mexico: The Role of Off-Farm Activities". *World Development,* 29(3): 467-480.

de Janvry, A. and E. Sadoulet (2004). "Progress in the Modelling of Rural Households' Behaviour under Market Failures". in *Poverty, Inequality and Development*, essays in Honour of Erik Thorbecke. Ed. by A. de Janvry and Ravi Kanbur. Kluwer.

de Luna, X. and P. Johansson (2001). "Testing Exogeneity under Distributional Misspecification". IFAU-Office of Labour Market Policy Evaluation. Working Paper 9.

Deaton, A., J. Ruiz-Castillo and D. Thomas (1989). "The Influence of Household Composition on Household Consumption Patterns: Theory and Spanish Evidence". *Journal of Political Economy,* 97(1): 179-200

Deaton, A. (1988). "The Allocation of Goods within the Household: Adults, Children and Gender". Population and Human Resources Department, World Bank. LSMS Working Paper no. 39.

Deaton, A. (1997). "The Analysis of Household Surveys. A Microeconometric Approach to Development Policy". World Bank.

Deaton, A. and J. Muellbauer (1980). "An Almost Ideal Demand System". *The American Economic Review*, 70:312-326.

Deere, C.D. (1986). "Rural Women and Agrarian Reform in Peru, Chile and Cuba". in *Women and Change in Latin America.* June Nash and Helen Safa (eds.) South Hadley, Mass.

Deere, C.D. and M. Leon (1982). "Women in Agriculture: Peasant Production and Proletarianization in Three Andean Regions". Geneva. ILO and Income Diversification in Colombia". *World Development,* 29(3):455-465.

Deininger, Klaus and Pedro Olinto (2001). *Rural Nonfarm Employment and Income Diversification in Colombia.* World Development vol. 29. No. 3. 455-465.

Deolalikar, A.B. and W.P.M. Vijverberg (1987). "A Test of Heterogeneity of Family and Hired Labour in Asian Agriculture". *Oxford Bulletin of Statistics*

Dercon, S. (1998). "Wealth, Risk and Activity Choice: Cattle in Western Tanzania". *Journal of Development Economics,* 55, 1:1-42

Dercon, S. and P. Krishnan (1996). "Income Portfolios in Rural Ethiopia and Tanzania: Choices and Constraints". *Journal of Development Studies,* 32(6):850-875.

Dercon, S. and P. Krishnan (1997). "In Sickness and in Health...Risk Sharing within Households in Rural Ethiopia". Centre for the Study of African Economies. WPS 97/12.

Donni, O. (2001). "Collective Labour Supply: Theory and Application". Laval, IZA Discussion paper 1506.

Donni, O. (2004). "A Collective Model of Household Behaviour with Private and Public Good: Theory and some Evidence from US Data". DELTA Working paper 2004-04.

Duflo E. and C. Udry. (2001). *Risk and Intrahousehold Resource Allocation in Cote d'Ivoire.* Preliminary Working Paper.

Efstratoglou-Todoulou, S. (1990). "Pluriactivity in Different Socio-Economic Contexts: a Test of the Push-Pull Hypothesis in Greek Farming". *Journal of Rural Studies,* 6(4): 407-413

Elbers, C. and P. Lanjouw (2001). Intersectoral Transfer, Growth and Inequality in Rural Ecuador. *World Development,* 29(3): 481-496.

Ellis, F. (1998). "Household Strategies and Rural Livelihood Diversification". *Journal of Development Studies,* 35(1):1-38.

Ellis, F. (2001). "Rural Livelihood Diversity in Developing Countries: Analysis, Policy Methods". Monologue. University of East Anglia.

Escobal, J. (2001). "The Determinants of Nonfarm Income Diversification in Rural Peru". *World Development*, 29(3): 497-508.

Eswaran, M. and A. Kotwal (1989). "Credit as Insurance in Agrarian Economies". *Journal of Development Economics,* 31: 37-53.

Evans, H, P. Ngau (1992) "Rural-urban relations: household income, diversification and agricultural productivity". *Development and Change,* 22:519-45.

Fafchamps Marcel and Agnes R. Quisumbing (1997) *Human Capital, Productivity and Labour Allocation in Rural Pakistan.* Working Paper. Stanford.

Fafchamps, M. and A.R. Quisumbing (2000). Social Roles, Human Capital and the Intra-Household Division of Labour: Evidence from Pakistan. 2000. Unpublished Manuscript.

Farrell, M.J. (1957). The Measurement of Production Efficiency. *Journal of the Royal Statistical Society Series A* 120 253-290

Feder, G. (1985). "The Relationship between Farm Size and Farm Productivity. The Role of Family Labour, Supervision and Credit Constraint". *Journal of Development Economics,* 18: 297-313.

Feder, G., L.J. Lau, J.Y.F. Lin and X. Luo. (1990). "The Relationship between Credit and Productivity in Chinese Agriculture: A Microeconomic Model of Disequilibrium." *American Journal of Agricultural Economics,* 1151-1157

Fernandez, M.E. (1988). "Technological Domains of Women in Mixed farming Systems of Andean Peasant Communities" in *Gender Issues in Farming Systems Research and Extension*. Susan V. Poats, M. Schmink and A. Spring (eds.) Boulder Colo. Westview Press. pp 61-72.

Ferreira F. and P. Lanjouw (2001) "Rural Poverty and Non-Farm Employment in Brazil". *World Development* 29(3): 509-528.

Filho C.A. and J.B. de Souza Ferreira Filho (2003) *Sources of Small Family Farm Production Inefficiency, reconcavo rgeion, Bahia, Brazil.* Revista de Economia e Sociologia Rural, vol 41, no. 1, Brasilia. Jan/March 2003.

Findeis, J. and D. Lass (1994). "Labour Decisions by Agricultural Households: Interrelationships Between Off-Farm Labour Supply and Hired Labour Demand". Population Research Institute, Pennsylvania State University.

Flora, C.B. and B. Santos (1986). "Women in Farming Systems in Latin America" in *Women and Change in Latin America*. J. Nash and H. Safa (eds.) South Hadley, Mass.

Fortin, B. and G. Lacroix (1997). "A Test of the Unitary and Collective Models of Household Labour Supply". *The Economic Journal*, 107(443): 933-955.

Fraser, I. and D. Cordina (1999). "An Application of DEA to Irrigated Dairy Farms". *Agricultural Systems,* 59: 267-82

Fraser, I. and Graham M. (2005). "Efficiency Measure of Australian Dairy Farms: National and Regional Performance". *Australian Agri-business Review,* 13(7) web-based.

Furtan, W., G. van Kooten, S.Thompson (1985). "The Estimation of Off-farm Labour Supply Functions in Saskatchewan". Journal of *Agricultural Economics*, 36: 211-9.

Gerfin, M. and G. Wanzenreid (2003). "Does it Matter Who Earns the Bread? A Test of the Income Pooling Hypothesis with Swiss Data". mimeo. Dept. of Economics, Bern.

Golan A., Judge G. and D. Miller (1996). "Maximum Entropy Economterics: Robust Estimation with Limited Data". Wiley.

Golan, A., J. Perloff and Z. Shen (2001). "Estimating a Demand System with Non-Negativity Constraints: Mexican Meat Demand". *Review of Economics and Statistics* 83(3) 541-550.

Goldstein, M. (2004). "Intrahousehold Efficiency and Individual Insurance in Ghana". Development Economics Discussion Paper Series. No. 38. LSE.

Gould, B. W., and Saupe, W. E. (1989). "Off-Farm Labor Market Entry and Exit". *American Journal of Agricultural Economics*, 71(4):960-969

Grimard, F. (2000). "Rural Labour Markets, Household Composition and Rainfall in the Cote dIvoire". *Review of Development Economics*, 4(1): 70-86

Hall, R. E. (1973). Wages, Income and Hours of Work in the US Labour Force. In *Income Maintenance and Labour Supply* G.C. Cain and H.W. Watts (eds.) (New York Academic press): 102-162.

Hamilton, S. Asturias L. and B. Tevalan (2001). "Gender and Agricultual Commercialization in Ecuador and Guatemala". IPM CRSP Working Paper 01-1

Hart, G.P. (1986). "Power, Labour and Livelihood: Process of Change in Rural Java". Berkeley: University of California Press.

Hartley, M. (1978). "Estimating Mixtures of Normal Distributions and Switching Regressions". *Journal of the American Statistical Association,* 73(364): 747-48.

Hausman, J.A. (1980). "The Effects of Wages, Taxes and Fixed Costs on Women's Labour Force Participation". *Journal of Public Economics* 14:161-194.

Hausman, J. A. (1981). "Labour Supply". In *How Taxes Affect Economic Behaviour,* H. Aaron and J. Pechman (eds.). The Brookings Institution.

Hearn, D, K. MacNamara, L. Gunter (1996). "Local Economic Structure and Off-farm Labour Earnings of Farm Operators and Spouses". *Journal of Agricultural Economics,* 47:28-36.

Heckman, J. J. and T.E. McCurdy (1980). "A Life Cycle Model of Female Labour Supply". *Review of Economic Studies*, 47: 47-74.

Heien, D. and C. Wessells. (1990). "Demand System Estimation with Microdata: A Censored Regression approach". *Journal of Business and Economic Statistics*, 8: 365-371.

Henning, C.H.C.A. and A. Henningsen (2005). "Modelling Price Response of Farm Households under Imperfect Labour Markets: A Farm Household Approach to Family Farms in Poland". Paper presented in Wye. 94th EAAE Seminar. Session 2a.

Huffman, W. (2004). "Discussion: Off-farm employment, government policy, and the structure of agriculture - An international perspective". *American Journal of Agricultural Economics,* 86(3): 737-739.

Huffman, W. E. and Lange, M. D. (1989) "Off-Farm Work Decisions of Husbands And Wives: Joint Decision Making". *Review of Economics and Statistics,* 79:471-480

Jacoby, H. (1993). "Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra". *Review of Economics Studies,* 60:903-21.

Jaforullah, M. and J. Whiteman (1999). "Scale Efficiency in New Zealand Dairy Industry: a non-Parametric Approach". *The Australian Journal of Agricultural and Resource Economics,* 43 (3), 523-41

Jha R., Chitkara P. and Gupta S. (2000) *Productivity, Technical and Allocative Efficiency and Farm Size in Wheat Farming in India*, a DEA Approach. Applied Economics Letters 7(1): 1-5.

Kalirajan, K. and R. Shand (1999). "Frontier Production Functions and Technical Efficiency Measure". *Journal of Economic Surveys*, 13 (2): 148-72

Kao, C., L. Lee and M. Pitt. (2001). "Simulated Maximum Likelihood Estimation of Linear Expenditure System with Binding Non-negativity Constraints". *Annals of Economics and Finance,* 203-223.

Keane M. (1992) ."A Note on Identification in the Multinomial Probit Model". *Journal of Business and & Economic Statistics* 10(2) 193-200.

Keeley, M.C., P. Robins, R. Spiegelman and R. West (1978). "The Estimation of Labour Supply Models using Experimental Data". *American Economic Review* 68. 873-887.

Kevane, M. (1996). "Agrarian Structure and Agricultural Practice: Typology and Application to Western Sudan". *American Journal of Agricultural Economics*, 78: 236-245.

Key, N., E. Sadoulet and A. de Janvry. (2000). "Transaction Costs and Agricultural Household Supply Response." *American Journal of Agricultural Economics,* 82(2):245-259.

Keeney M. (2000) "The Economic Consequences of Direct Payments for Irish Agriculture". PhD thesis. Trinity College Dublin.

Kimhi, A and Lee, M-J. (1996). "Off-Farm Work Decisions of Farm Couples: Estimating Structural Simultaneous Equations With Ordered Categorical Dependent Variables". *American Journal of Agricultural Economics,* 78:687 – 698

Kimhi, A. (1994). "Quasi Maximum Likelihood Estimation of Multivariate Probit Models: Farm Couples' Labour Participation". *American Journal of Agricultural Economics,* 76:828-35.

Kochar, A. (1999). "Smoothing Consumption by Smoothing Income: Hours of Work Responses to Idiosyncratic Agricultural Shocks in Rural India". *Review of Economics and Statistics* 81: 50-61.

Koolwall, G. and R. Ray (2002). "Estimating the Endogenously Determined balance of Power and its Impact on Expenditure patterns: Evidence from Nepal". Washington D.C. World Bank. WPS2814.

Kooreman, P. and A. Kapteyn (1987). "A Disaggregated Analysis of the allocation of Time with a Household". *Journal of Political Economy*, 95(2): 223-249.

Kuroda, Y. and P. Yotopoulos (1978). "A Microeconomic Analysis of Production Behavior of the Farm Household in Japan :A Profit Function Approach" *The Economic Review (Japan),* 29:115- 129.

Lamb, R.L. (2003). "Inverse Productivity, Land Quality, Labour Markets and Measurement Errors". *Journal of Development Economics,* 71(1): 71-96.

Lambert, S. and T. Magnac (1994). Measurement of Implicit Prices of Family Labour in Agriculture: An Application to Cote d'Ivoire. In Caillavet, Gyomard and Lifran (eds.) A*gricultural Household Modelling and Family Economics*. Amsterdam:Elsevier.

Lancaster, G., P. Maitra and R. Ray (2003). "Endogenous Power, Household Expenditure Patters and New Tests of Gender Bias: Evidence from India". Preliminary Working Paper. Tasmania.

Lanjouw, J.O. and P. Lanjouw (2001). "The Rural Non-Farm Sector: Issues and Evidence from Developing Countries". *Agricultural Economics,* 26: 1-23.

Lanjouw, P. (1998). "Poverty and the Non-Farm Economy in Mexico's Ejidos: 1994-1997". Draft Paper. Vrije University of Amsterdam (VUA).

Lanjouw, P. (2001). "Nonfarm Employment and Poverty in Rural El Salvador". *World Development,* 29(3), 529-547.

Lanjouw, P., J. Quizon and R. Sparrow (2001). "Non-agricultural Earnings in Peri-urban Areas of Tanzania: evidence from Household Survey Data". *Food Policy,* 26:385-403.

Lass, D. and C. Gempeshaw (1992). "The Supply of off-farm labour: a random coefficients approach". *American Journal of Agricultural Economics,* 74:400-11.

Lau, L.J., P. Yotopoulos, E. Chou and W. Lin (1978). "The Microeconomics of Distribution: A Simulation of the Farm Economy." *Journal of Policy Modelling,* 3:175-206.

Lazlo, S. (2001). "Self-Employment Earnings: An Empirical Analysis of Non-Farm Activities in Rural Peru". Dept.of Economics Ryerson University and Dept. of Economics, University of Toronto.

Lazlo, S. (2002). "Education, Labour Supply and Market Development in Rural Peru". Working Paper. Ryerson/Toronto.

Lee, M. (1998). "Off-farm Labour Supply and Various Related Aspects of Resource Allocation by Agricultural Households". PhD Dissertation, Gottingen University.

Lockheed, M.E., D.T. Jamison and L.J. Lau (1980). "Farmer Education and Farm Efficiency: A Survey". *Economic Development and Cultural Change*, 29(1), 37-76.

Lofgren, H. and S. Robinson (1999). "To Trade or not to Trade: Non-Separable Farm Household Models in Partial and General Equilibrium". *International Food Policy Research Institute.* TMD Discussion paper 37.

Lopez, R. (1984). "Estimating Labor Supply and Production Decisions of Self-Employed Farm Producers". *European Economic Review, 24:*61 -82

Lundberg, S (1988). "Labor Supply of Husband and Wives: A Simultaneous Equations Approach". *Review of Economics and Statistics* 70(2): 224-235

Lundberg, S., R. Pollak and T. Wales (1997). "Do Husbands and Wives Pool their Resources"? *Journal of Human Resources*, 32(3): 463-480.

Malchow-Moller, N. and M. Svarer (2002). "Estimation of the Multinomial Logit Model with Random Effects". Centre for Economics and Business Research, Copenhagen. Discussion Paper no. 2002-16.

Manser, M. and M. Brown (1980). "Marriage and Household Decision Making: A Bargaining Analysis". *International Economic Review,* 21:31-44.

McCurdy, T. E. (1981). "An Empirical Model of Labour Supply in a Life-Cycle Setting". *Journal of Political Economy,* 89:1059-1085.

McElroy, M.B. and Horney, M.J. (1981). "Nash-Bargaining Household Decisions: Toward a Generalization of the Theory of Demand". *International Economic Review*, 22:333 – 49

Mendoza, L. (2002). "Testing for the Existence of Bargaining in Rural Households. A Study of Decisions on Labour Market Participation in the Cordillera Region of the Phillipines". Leuven; part of dissertation under L. Berlage.

Mishra, A., Goodwin B. (1997). "Farm income variability and the supply of off-farm labour". *American Journal of Agricultural Economics,* 79:880-7.

Mittlehammer, R., G. Judge, D. Miller. (2000). "Econometric Foundations". Cambridge University Press.

Morduch, J. (1995). "Income Smoothing and Consumption Smoothing". *Journal of Economic Perspectives*, 9(3):103-114.

Murrugarra, E. (1998). "Labour Demand and Incomplete Markets: Testing for Nonseparation on Peruvian Rural Households". UCLA. PhD. Dissertation.

Nash, J. (1986). "A Decade of Research on Women in Latin America" in *Women and Change in Latin America"*. J. Nash and H. Safa (eds.). South Hadley, Mass.

Newman, J. and P.J. Gertler (1994). "Family Productivity, Labour Supply and Welfare in a Low Income Country". *Journal of Human Resources*, 29: 989-1026.

Olfert, M. (1992). "Non-farm Employment as a Response to Underemployment in Agriculture". *Canadian Journal of Agricultural Economics,* 40:443-58.

Olfert, M., Taylor J., Stabler J. (1993). "Non-farm Labour Market Participation of Farm Women". *Canadian Journal of Agricultural Economics,* 41:81-95.

Onamo, S. (1998). "Farm-to-Market Transaction Costs and Specialization in Small-Scale Agriculture Explorations with a non-Separable Household Model". *Journal of Development Studies* 35(2), 152-163.

Pascual, U. and E. Barbier (2001). "Determinants of Labour Supply by Shifting Cultivator Households. The Effect of Land Quality and Poverty". Paper Presented at the EAERE 2001 11th Annual Conference at Southampton.

Paulson, A. (2000). "Insurance Motives for Migration". Evidence from Thailand. Draft. Priceton University". Kellogg Graduate School of Management. Working Paper.

Paxson, C. (2000). "Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand". in Bardham P. and Udry C. (eds). *Readings in Development Economic*, MIT Press.

Phillips, J.M. (1987). "A Comment on Farmer Education and Farm Efficiency: A Survey". *Economic Development and Cultural Change*, 35(3): 637-644. April. 1987.

Phipps, S. and P. Burton (1992). "What's Mine is Yours? The Influence of Male and Female Income on Patterns of Household Expenditure". Working Paper. Halifax. N.S. Dalhousie Univ. 92-12.

Pinheiro A.E. and B.E. Bravo-Ureta (1997) *Technical, Economic and Allocative Efficiency in Peasant Farming: Evidence from the Dominican Republic.* The Developing Economies XXXV-1. March. Pp 48-67

Pitt, M.M. and M. Rosenzweig (1986). "Agricultural Prices, Food Consumption and the Health and Productivity of Indonesian Farmers". In I. Singh, L. Squire and J. Strauss eds. *Agricultural Household Models*. Baltimore John Hopkins Press.

Rangel, M.A. (2003). "Alimony Rights and Intrahousehold Bargaining: Evidence from Brazil". CCPR-004-03

Rapoport, B., Sofer C. and Solaz A. (2003). "Household Production in a Collective Model. Some New Results". TEAM. Paris presented at ASSA meeting, San Diego, 2004.

Ray S.C. and Bhadra D. (1993) "Nonparametric Tests of Cost Minimizing Behaviour: a Study of Indian Farms". *American Journal of Agricultural Economics*, v. 75, n 4, p 990-999.

Reardon, T. and M. Mercvado-Peters (1993). "Self-financing of Rural Household Cash Expenditures in Burkina Faso: the case of net cereal buyers'". In Cuevas and Benoit-Cattion (eds.), *Finance and Rural Development in West-Africa*, seminar, CIRAD.

Reardon, T., E. Crawford and V. Kelly (1994). "Links between Nonfarm Income and Farm Investment in African Households: adding the Capital Market Perspective". *American Journal of Agricultural Economics*, 76(5): 1172-1176.

Reardon, T. (1997). "Using Evidence of Household Income Diversification to Inform Study of the Rural Nonfarm Labour Market in Africa". *World Development*, 25, 735-747.

Reardon Thomas, Julio Berdegue and German Escobar. (2001) Rural Nonfarm Employment and Incomes in Latin America: Overview and Policy Implications. *World Development* vol. 29. (3). pp 395-409.

Robles, M. (1997). "Pobreza e Inequidad en el Peru, 1996". Programa Mecovi-Peru. BID. BM.CEPAL.

Rose, E. (1994). "Consumption Smoothing and Excess Female Mortality in Rural India". Draft. University of Washington.

Rosenzweig, M.R. (1980). "Neoclassical Theory and the Optimizing Peasant: An Econometric Analysis of Market Family Labor Supply in a Developing Country," *The Quarterly Journal of Economics*, 94(1):31-55.

Rosenzweig, M.R. (1988). "Risk, Implicit Contracts and the Family in Rural Areas of Low Income Countries". *Economic Journal*. December, 1988. 98. 1148-70.

Rosenzweig, M.R. (1989). "Credit Market Constraints, Consumption Smoothing and the Accumulation of Durable Production Assets in Low Income Countries: Investments in Bullocks in India". Manuscript. Minnesota.

Rosenzweig, M.R. and O. Stark (1989). "Consumption Smoothing, Migration and Marriage: Evidence from Rural India". *Journal of Political Economy*, 97:4. 905-926.

Rozelle, S., J.E. Taylor, and A. de Brauw (1999). "Migration, Remittances, and Productivity in China." *American Economic Review*, 89(2), 287-91.

Ruben, R. and M. van den Berg (2001). "Nonfarm Employment and Poverty Alleviation of Rural Farm Households in Honduras". *World Development,* 29(3): 549-560.

Sadoulet, E., A. de Janvry and C. Benjamin. (1998). "Household Behaviour with Imperfect Labour Markets". *Industrial Relations*, 37(1): 85-108.

Samuelson, P. A. (1956). "Social Indifference Curves" *Quarterly Journal of Economics* vol. LXX, 1 – 22.

Savadogo, K., T. Reardon and K. Pietola (1998). "Adoption of Improved Land-use Technologies to Increase Food Security in Burkina Faso: Relating Animal Traction, Productivity and Non-Farm Income". *Agricultural Systems*, 58 (3): 441-464.

Schady, N.R. (2000) "Picking the Poor: Indicators for Geographic Targeting in Peru". World Bank Policy Research Working Papers number 2477.

Schady, N.R. (2002). "The (Positive) Effect of Macroeconomic Crises on the Schooling and Employment Decisions of Children in a Middle-Income Country". World Bank. Policy Research Paper number 2762.

Schultz, T.P. (1990). "Testing the Neoclassical Model of Family Labor Supply and Fertility," *Journal of Human Resources* 25(4):599-634.

Sen, A. (1966). "Peasants and Dualism with or without Surplus Labour". *Journal of Political Economy,* 74: 425-450.

Shafiq, M. and T. Rehman (2000). "The Extent of Resource Use Inefficiencies in Cotton Production in Pakistan's Punjab: An application of DEA". *Agricultural Economics,* 22: 321-330.

Sharam, K. R., P. Lueng, H.M. Zaleski (1999). "Technical, Allocative and Economic Efficiencies in Swine Production In Hawaii: A Comparison of Parametric and Nonparametric Approaches". *Agricultural Economics,* 20: 23-35.

Shonkweiler, J.S. and S.T. Yen. (1999). "Two-step Estimation of a Censored System of Equations". *American Journal of Agricultural Economics*, 81(4): 972-82.

Simar, L. and P.W. Wilson (2000). "Statistical Inference in Nonparametric Frontier Models., the State of the Art". *Journal of Productivity Analysis*, 13, 49-78

Singh, I, Squire L, Strauss J. (1986). "Agricultural Household Models". Baltimore and London: John Hopkins University Press.

Skoufias, E. (1994). "Using Shadow Wages to Estimate Labour Supply of Agricultural Households". *American Journal of Agricultural Economics,* 76:215-27.

Smith, L.C. and J.P. Chavas (1999). "Supply Response of West African Agricultural Households. Implications of Household Preference Heterogeneity". FCND Discussion Paper 69.

Stark, O. (1991). "The Migration of Labor". Cambridge, MA: Basil Blackwell.

Strauss, J. and D. Thomas (1995). "Human Resources, Empirical Modelling of Household Family Decisions". *Handbook of Development Economics, vol. 3.* J. Behrman and T.N. Srinivasan (eds.). New York. North Holland.

Sumner D. (1982). "The Off-Farm Labour Supply of Farmers". *American Journal of Agricultural Economics,* 64:499-509.

Swamy, G. (1981). "International Migrant Workers Remittances: Issues and prospects". World Bank, Washington D.C. Staff Working Paper No. 481.

Taylor, J.E and I, Adelman (2002). "Agricultural Household Models: Genesis, Evolution and Extensions". *Review of Economics of the Household,* 1(1).

Thomas, D. (1990). "Intra-Household Resource Allocation: An Inferential Approach". *The Journal of Human Resources* 25, 635-663.

Thomas, D. and C.Chen (1994). "Income Shares and Shares of Income: Empirical tests of Models of Household Resource Allocations". Working Paper Series. No. 94-08. Rand Corp.

Tokle, J. and W. Huffman (1991). "Local Economic Conditions and Wage Labor Decisions of Farm and Rural Nonfarm Couples". *American Journal of Agricultural Economics* pp652 – 670

Townsend R. F., J. Kirten, N. Vink (1998). "Farm Size Productivity and Returns to Scale in Agriculture Revisited: A Case Study of Wine Producers in South Africa". *Agricultural Economics,* 19: 175-180.

Townsend, O. (1994). "Risk and Insurance in Village India". *Econometrica,* 62: 539-92.

Udry, C. (1996). "Gender, Agricultural Production and the Theory of the Household". *Journal of Political Economy*, 104(5): 1010-1046.

Udry, C. and T.G. Conley (2004). "Social Networks in Ghana". Centre Discussion Paper 888. Economic Growth Centre. Yale University.

Vakis, R., E. Sadoulet, A. de Janvry and C. Cafiero (2004). "Testing for Separability in Household Models with Heterogeneous Behaviour: A Mixture Model Approach". CUDARE Working papers. Berkeley.

Van der Ploeg, J.D. (1990). "Sistemas de conocimiento, metafora y campo de interaccion el caso del cultivo de la papa en el altiplano peruano". *Agricultura y Sociedad* 56 (julio/septiembre) 143-166.

Vermeulen F. (2002). "Collective Household Models: Principles and Main Results.". *Journal of Economic Surveys*, 2002. 16 (4). 533-564.

Vermeulen, F. (2005). "And the Winner Is....An Empirical Evaluation of Unitary and Collective Labour Supply Models". *Empirical Economics*, 2005, 30(3). 711-734.

Weersink A. C.G. Tunvey and A. Godah (1990) "Decomposition of Technical Efficiency Measures for Ontario Dairy Farms". *Canadian Journal of Agricultural Economics*. 38(3) 439-456.

Weersink, A., Nicholson C, Weerhewa J. (1998). "Multiple job-holding among Dairy Farm Families in New York and Ontario". *Agricultural Economics,* 18:127-43.

Wiens, T. (1977). "Uncertainty and Factor Allocation in a Peasant Economy". *Oxford Economic Papers,* 29: 48-60.

Woldehanna, T., Lansink A., Peerlings J. (2000). "Off-farm work decisions on Dutch cash crop farms and the 1992 and Agenda 2000 CAP Reforms". *Agricultural Economics,* 22:163-71.

Woldehanna, T. and Oskam A. (2001). "Income Diversification and Entry Barriers: Evidence from the Tigray region of Northern Ethiopia". *Food Policy,* 26. 351-365.

Xue, M. and P.T. Harker (1999). "Overcoming the Inherent Dependency of DEA Efficiency Scores: A Bootstrap Approach". Wharton Financial Institutions Centre. Working Paper 99-17.

Yang, D. T. (2000). "Education and Allocative Efficiency: Household Income Growth during Rural Reforms in China". Working Paper 00-17. Department of Economics. Duke University.

Yen S.T., B.H. Lin and D.M. Smallwood. (2003) *Quasi and Simulated Approaches to Censored Demand Systems. Food Consumption by Food Stamp Recipients in the United States*. American Journal of Agricultural Economics. 85(2) may. 458-478.

Zaibet, L. and P.S. Dharmapala (1999). "Efficiency of Government-Supported Horticulture: The Case of Oman". *Agricultural Systems,* 62: 159-68.

Zamora, B. (2003). "The Spanish Sharing Rule". Alicante University. Economics Working paper Archive 0302005.

Zhang, J. and K. Fong (2001). "The Identification of Unobservable Independent and Spousal Leisure". *Journal of Political Economy*, 109(1). 191-202

Zhang, Y. and R. Bartels (1998). "The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand". *Journal of Productivity Analysis,* 9: 187-204.

Zhiyang Jia. (2005) "Labour Supply of Retiring Couples and Heterogeneity in Household Decision-Making Structure". Statistics, Norway. Research Department Working Paper.

Appendices

Chapter 1: **190**
Appendix 1: Collective model in farm context 190
Chapter 2 .... **195**
Chapter 3 .... **196**
Parameters of AIDS models and basis for 3-decision maker test 196
Chapter 4 .... **201**
Chapter 4: Appendix 1: Tables 201
Multivariate probit models 201
Individualised multinomial logit models 201
Instrumental variable Models for Table 4.1 204
Chapter 4; Appendix 2. Variables used in the reduced form participation and hours models (with reference also to the models of the other chapters) 205
Chapter 4; Appendix 3. Reduced form hours estimations 212
Introduction 212
Labour market considerations and interpretation of reduced form coefficients 214
Hours estimations 216
Discussion of hours results 222
Chapter 5 .... **224**
Chapter 6 .... **232**

Table 2-A 1: Cluster loadings 195

Table 3-A 1: Entropy model (*100) 196
Table 3-A 2: Two-step model with selection variable 196
Table 3-A 3: IV model parameters 197

Table 4-A 1: Correlations from the multivariate probit models 201
Table 4-A 2: Individualised Multinomial Logit Models 201
Table 4-A 3: Instrumental variable models for table 4.1 204
Table 4-A 4: Household Selection Model 217
Table 4-A 5: Hours equations based on the bivariate probit selection process 218
Table 4-A 6: Off-farm hours (in 4 categories) 219

Table 5-A 1: Test for linearity of self-employed returns 224
Table 5-A 2: Test of separability in off-farm work for households with just one wage earner 225
Table 5-A 3: Multinomial logit for 8 individual regimes 225
Table 5-A 4: Heterogeneous households in constrained and unconstrained regimes 227
Table 5-A 5: Summary Statistics for constrained and unconstrained households 228
Table 5-A-6: Multinomial participation model for table 5-3 228

Table 6-A 1: CLAD model using off-farm earnings, not hours 232

Chapter 1:

Appendix 1: Collective model in farm context[39]

In this appendix the application of the Collective model to a farm household context is discussed. Let p be the price vector for farm goods. Then p*g(t, v) is farm revenue, where g is a production function, v are farm productivity characteristics, and t is total household farm labour (for now, hiring in labour is assumed not to occur). If markets are smooth, then farm profits, abstracting from inputs, are: $pg' = pg - \sum_{i=1} w_i t_i$.

Given that this is fundamentally the same set-up as a household production model with marketable goods, the results for household production shown in Apps and Rees (1995), Chiappori (1997), and extended in Rapoport *et al.* (2003), hold and the actual sharing rule is again recoverable up to an additive constant. Specifically, for the simplest (least data intensive) form of the model, if there is separability of consumption and production, and if returns to scale to on-farm production are constant or falling, then farm profits are exogenous, so the two step logic of the sharing rule programme continues to operate (farm profits and non-labour income are shared, and then labour supply and consumption are decided). Total available personal income (the individual's budget constraint) becomes $w_i(h_i + t_i) + \phi_i + pg_i'$, or $w_i(h_i + t_i) + \psi_i$, where ψ_i is the sum of farm profits and own share of non-labour income. The new sharing rule ψ_i is a function of wages, unearned income, the distribution factor(s), the price vector for farm goods, and wages for the operator and spouse and z (the personal and household characteristics affecting preferences): $\psi_1 = \psi_1(w_1, w_2, y, z, s, p)$, and Marshallian labour supply functions (for total labour time: $h_i + t_i$) are changed accordingly: $h_i + t_i = H_i(w_i, \psi_i(w_{i,} w_j, y, z, s, p), z)$. Thus – if prices are everywhere exogenous and equal – the only extra data needed compared to the usual Collective Model is individual time spent in farm work (or, to be precise, total individual time spent in farm and off-farm work). It is important to note that neither individual on-farm productivities (v_i) nor farm level productivity indicators (v) play any part in this allocation. The assumption is that only market characteristics

[39] What follows is largely an attempt to apply the work on household production in Chiappori (1997), Apps (2003), and Rapoport, Sofer and Solaz (2003) to the case of farm production.

play any role in increasing or decreasing individual bargaining strength. This would seem to be a fairly unrealistic assumption, and it can be improved on.

If we have further information (specifically, on farm characteristics, output prices and on-farm labour time allocation), and if we assume some form of the Hicks composite good theorem or else separability in production if there is more than one farm output (not such a strong assumption for some farm tasks, but fairly strong perhaps for others), as well as a common (or separately identifiable) production function across households, then it becomes possible to separate out ϕ_i from $pg_i^{'}$. This is important if we want to separate out the shares of non-labour income from shares of farm profits. When we do this, we get the following decomposition: $\phi_i(w_1, w_2, y, z, p; v, v_i, z) = \psi_i(w_1, w_2, y, s, p; v, v_i, z) - pg_i^{'}(w_{1,} w_2, v_i, v, z)$. Share of farm profits now depend on prices, on individual productivities, on v and z, and on the opportunity costs of working on the farm (the two wage rates). This means that, given the data, we can recover up to an additive constant the individual shares both of farm profits and of non-labour income. For any variable x, where x= w_1, w_2, y, v, v_i, p, $\partial\psi_i / \partial x_i$ can be broken down into: $(\partial pg_i^{'} / \partial x) + (\partial\phi_i / \partial x)$. An increase in operator's off-farm wages, for example, because of the increase in bargaining power it leads to, affects both his share of farm profits and his share of labour and non-labour income. The predicted income effects on his labour supply must be different therefore from those given by the Unitary model (which assumes constant shares); likewise, the effects of his wage increase on his partner's off-farm labour supply will almost certainly differ from the effect predicted effects under the assumptions of the Unitary model. Some of these differences in predictions between the Collective and the Unitary models are now briefly explored.

With regard to differences in interpretation of and type of results generated from the Unitary Model, one important difference between the Unitary model and the general version of the Collective Model is that, unlike with the Unitary Model, the matrix of compensated effects (the Collective version of the Slutsky matrix) of partner's wage changes on own labour supply is not symmetric in the Collective Model (where both partners work off-farm), so the compensated operator's supply response to a change in spousal wages need not equal the spouse's compensated response to a change in operator's wages. The reason for the lack of symmetry is the change in individual bargaining positions that results from the wage changes. The wage changes affect the sharing rule, even holding income constant. Compensated changes are thus a sum of three effects – the familiar substitution effect, the effect of the wage change on the sharing rule, and the effect of the change in shares on actual behaviour. Only if the sharing rule is unaffected is the Slutsky matrix symmetric, in which case a restricted form of the Unitary model holds (Gorman weak separability).[40]

Returning to the default 'caring' version of the model, in the Unitary Model, where both partner's work off-farm, if operator and spouse are gross substitutes in farm production and Hicksian substitutes in home time, then the spouse's shadow wage rises as operator's wages rise, and thus her off-farm labour supply falls (Lee, Chapter 6). The size, and even the direction, of these effects are mediated in the Collective Model via the sharing rule. Assuming the effect on the share is positive ($\phi^2_{w_1} > 0$), and always that leisure is normal, then an increase in operator's wages leads to a rise in the spouse's income, her budget constraint is loosened, and – for a given level of on-farm labour time – her off-farm labour supply falls. On-farm adjustment itself then depends on the two productivities and the degree of their substitutability or complementarity, and can lead to either an increase or decrease in off-farm work and leisure. A negative direct effect on the share is also conceivable ($\phi^2_{w_1} < 0$), though very unlikely (it means that the operator appropriates more than 100% of any own wage gain). If such an effect were to occur, it would lead to a tightening of the spouse's budget constraint and – for a given level of on-farm work – an increase in off-farm labour. Again, the final allocation would depend on the interplay of on-farm productivities.

Turning to output prices, in the Unitary Model an exogenous increase in the price of farm outputs affects off-farm labour as follows (again, with both partners assumed to be working on and off the farm): returns to farm work rise, and so on-farm work rises until the respective marginal revenue products match wages. Participation in off-farm

[40] The Collective model does NOT generally nest the Unitary one: both can be seen as restricted versions of more general models. However, weakly separable forms of the Unitary model are nested by the Collective model. Also, with regard to the Slutsky matrix, Browning and Chiappori (1998) show that the particular form that the estimated adjusted Slutsky matrix takes can be used to test the Collective Model vis a vis the Unitary Model if preferences are general (allowing for externalities in private consumption), as long as there is a certain minimum of goods **with price variation** that can be used in testing (5, including leisure, in the case of a household with 2 decision makers). This test is used and explained further in section 3.3. A further set of related tests are possible if there are one or more distribution factors.

work is less likely for both partners, and – for those continuing to participate – hours in off-farm work will fall: the relative size of the effects on each partner are determined by the (somehow agreed, or dictated) joint tastes for leisure of the partners. In the Collective Model, on the other hand, the off-farm labour supply effect, once again, is through the sharing rule as it applies now to farm profits: returns to farm work rise and time in farm work rises as before; depending on the sharing rule the budget constraint may loosen for both partners, or maybe only for one (although this would seem to be unlikely), so the direction and size of the off-farm participation and hours effects are determined first by the extent of this loosening (determined, itself, for each individual, by the sharing rule) and only then by individual tastes for leisure. The first effect – if it exists - is simply not analysable within the Unitary framework (and, neither, strictly speaking, is the second, since the individual's taste for leisure has a different meaning in the Unitary model than in the Collective model: in the Unitary Model each individual's taste for leisure simply generates an argument in the household utility function, and leisure in some sense is allocated by the household as a unit; this is not the case in the Collective Model, where taste for leisure is explicitly the individual's[41]).

Thus, for most empirical work, the essential difference between the two models is in their delineation of income effects. In the Unitary Model the strength of the income effects for each partner depends on the shadow values allocated by the household, and not on the source of income changes. In the Collective Model, individual tastes matter, the source of the new income matters, and, in the last version of the model, individual on-farm productivities also matter. Income sources and productivities matter in particular because they affect bargaining strength within the family.

Furthermore, given time use data on housework, and given marketability of the products of housework, it would also appear to be possible, using the programmes developed in Chiappori (1997) and Rapoport *et al.* (2003), to add household production to farm work in the estimation of off-farm labour supply. In this case, the aim would be to recover the sharing rule for wages, farm profits plus – together - non-labour income and the (private) fruits of housework. The latter two are unlikely to be easily differentiable, as the fruits of housework would need to be valued to do so.

[41] Whether taste for leisure and strength of bargaining power in helping to gain more leisure are separably estimable with any great degree of confidence would seem very unlikely. In the Collective Model as of now, they are not.

Rapoport *et al.*, using French time use data, find that results are much sharper if household production is included in the labour supply estimations. It would be interesting to find if this pattern generalizes - the distribution factor they use (local gender ratios), for example, is found to be a significant determinant of income shares when household production is included as part of the dependent variable (hours at work), and insignificant otherwise; similarly non-labour income affects female labour supply in the expected way if household production is included, but not otherwise.

Appendix 1: Cluster loadings

Table 2-A 1: Cluster loadings

	Cluster				
	1	2	3	4	5
Age	0.452	-0.416	-0.087	-0.498	0.052
Years of education	-0.155	0.299	-0.159	0.310	-0.200
Operator	0.434	-0.219	-0.158	-0.568	0.103
Spouse	0.058	-0.060	0.317	-0.271	0.161
Other male	-0.399	0.031	0.054	0.666	-0.164
Other female	-0.282	0.356	-0.221	0.465	-0.178
Rainforest region	-0.619	-0.619	1.518	-0.619	1.613
Sierra region	0.843	0.843	-1.185	-0.210	-1.185
Costal region	-0.400	-0.400	-0.277	1.103	-0.400
Number of children (<15 years)	-0.185	2.636	-0.358	-0.326	-0.358
Banana	-0.266	-0.266	3.753	-0.266	-0.266
Spanish speaking	-0.379	-0.279	0.206	0.371	0.253
ln Farm capital	0.445	-0.028	-0.824	0.163	-0.687
ln Farm size	-0.198	-0.590	0.699	-0.178	0.602
On-farm diversification	0.061	-0.119	0.473	-0.197	0.043
Proportion of crops sold	-0.182	-0.687	0.204	0.438	0.019

Chapter 3

Parameters of AIDS models

Table 3-A 1: Entropy model (*100)

	Rice	Sugar	Meat	Water	Soap	Oil
Price of rice (ln)	-12.252	-3.172	-1.721	0.243	0.043	-2.947
	[0.00]	[0.00]	[0.03]	[0.75]	[0.89]	[0.00]
Price of sugar (ln)	7.411	3.144	2.578	-2.131	-0.028	2.994
	[0.00]	[0.00]	[0.02]	[0.05]	[0.96]	[0.01]
Price of meat (ln)	0.410	0.909	0.790	0.262	-0.011	0.747
	[0.19]	[0.00]	[0.01]	[0.41]	[0.96]	[0.02]
Price of water (ln)	-1.261	-4.163	-1.952	-0.839	-0.180	0.616
	[0.22]	[0.00]	[0.06]	[0.42]	[0.84]	[0.55]
Price of soap (ln)	-3.838	2.334	4.211	0.894	0.068	0.412
	[0.00]	[0.02]	[0.00]	[0.37]	[0.94]	[0.68]
Price of oil (ln)	2.410	1.288	-1.020	0.633	0.069	0.787
	[0.00]	[0.07]	[0.16]	[0.38]	[0.92]	[0.27]
Education	-0.240	0.243	0.713	0.129	0.027	0.397
	[0.32]	[0.33]	[0.01]	[0.61]	[0.91]	[0.12]
Family size	3.546	1.493	-0.318	-1.144	0.019	-0.800
	[0.00]	[0.00]	[0.37]	[0.00]	[0.95]	[0.02]
Total family expenditure	-4.252	-1.755	0.757	0.211	-0.135	-0.830
	[0.00]	[0.00]	[0.01]	[0.46]	[0.64]	[0.00]
Spanish speaking	3.778	-0.137	0.055	-0.198	0.062	0.501
	[0.00]	[0.70]	[0.87]	[0.58]	[0.86]	[0.16]
Number of rooms in the house	0.592	0.061	0.614	0.049	0.012	0.046
	[0.00]	[0.55]	[0.00]	[0.63]	[0.90]	[0.66]
Intercept	56.196	2.206	-28.894	-3.017	-6.643	-11.942
	[0.20]	[0.88]	[0.50]	[0.93]	[0.87]	[0.78]

Table 3-A 2: Two-step model with selection variable

	Rice	**Soap**	**Sugar**	**Water**	**Meat**	**Oil**
Selection	-0.631	-0.005	-0.129	0.068	-0.153	-0.273
	[0.00]	[0.03]	[0.01]	[0.14]	[0.00]	[0.00]
Price of rice (ln)	-0.001	0.000	-0.035	0.008	0.141	0.029
	[0.97]	[0.64]	[0.00]	[0.76]	[0.00]	[0.07]
Price of sugar (ln)	-0.078	0.001	0.003	-0.091	-0.232	-0.116
	[0.05]	[0.26]	[0.87]	[0.02]	[0.00]	[0.00]
Price of meat (ln)	-0.024	0.000	-0.003	0.011	-0.004	-0.023
	[0.01]	[0.47]	[0.59]	[0.24]	[0.63]	[0.00]
Price of water (ln)	0.108	-0.002	-0.025	-0.044	0.159	0.086
	[0.00]	[0.01]	[0.13]	[0.30]	[0.00]	[0.00]
Price of soap (ln)	-0.055	0.001	0.043	0.034	-0.029	0.028
	[0.04]	[0.33]	[0.00]	[0.28]	[0.35]	[0.05]
Price of oil (ln)	0.038	0.001	0.020	0.058	-0.008	-0.003
	[0.04]	[0.13]	[0.03]	[0.10]	[0.71]	[0.76]
Education	-0.009	0.000	0.004	0.008	-0.023	0.001
	[0.20]	[0.06]	[0.23]	[0.47]	[0.02]	[0.73]
Family size	-0.029	0.000	0.012	-0.029	-0.038	-0.012
	[0.01]	[0.24]	[0.02]	[0.01]	[0.00]	[0.02]
Total family expenditure	0.010	-0.001	-0.014	0.008	0.058	0.003
	[0.25]	[0.00]	[0.00]	[0.32]	[0.00]	[0.54]
Spanish speaking	-0.035	0.001	-0.007	-0.008	-0.047	-0.010
	[0.01]	[0.01]	[0.14]	[0.49]	[0.00]	[0.08]
Number of rooms in the house	-0.003	0.000	-0.001	0.002	-0.012	-0.003
	[0.34]	[0.05]	[0.41]	[0.59]	[0.00]	[0.08]
Intercept	0.320	0.006	0.082	0.001	0.011	0.140
	[0.00]	[0.01]	[0.06]	[0.85]	[0.27]	[0.00]
Observations	499	499	499	499	499	499
R^2	0.2943	0.1783	0.133	0.0491	0.2152	0.1024

Table 3-A 3: IV model parameters

	Rice	Sugar	Meat	Water	Soap	Oil
Price of rice (ln)	0.095	0.045	-0.114	-0.044	0.005	0.050
Price of sugar (ln)	-0.309	-0.103	0.197	0.062	-0.009	-0.112
Price of meat (ln)	-0.037	-0.005	0.027	0.012	-0.001	-0.008
Price of water (ln)	0.128	0.008	-0.081	-0.038	0.001	0.057
Price of soap (ln)	-0.023	0.029	0.035	0.006	0.001	0.010
Price of oil (ln)	0.032	0.015	-0.016	0.003	0.001	0.012
Education	-0.025	-0.006	0.016	0.006	0.000	-0.004
Family size	-0.047	-0.015	0.031	0.005	-0.002	-0.036
Total family expenditure	0.170	0.059	-0.079	-0.039	0.004	0.065
Spanish speaking	-0.011	-0.019	0.021	0.008	-0.001	-0.012
Number of rooms in the house	-0.010	-0.005	0.013	0.004	0.000	-0.005
Intercept	-0.021	-0.123	0.081	0.198	0.037	-0.213

Basis for the result reported in Equation 3 – 6 (the three-decision-maker restriction):

Given that, as noted in the text, the matrix M, where $M = S - S'$ (and $S = \Sigma + uv'$), is an antisymmetric matrix and that a real antisymmetric matrix must have even rank, Browning and Chiappori (1998), for two decision maker households, show that M will have a rank of at most 2, whereas the symmetric Slutsky matrix Σ has a rank of zero. In the five good model estimated in their paper, the 4*4 M matrix (where the fifth good ensures adding up) would look as follows (they do not actually show it in the paper): $\begin{bmatrix} 0 & m_{12} & m_{13} & m_{14} \\ -m_{12} & 0 & m_{23} & m_{24} \\ -m_{13} & -m_{23} & 0 & m_{34} \\ -m_{14} & -m_{24} & -m_{34} & 0 \end{bmatrix}$.

If M is of rank 2 and the bottom two rows are linearly dependent on the top two rows, then m^3 (where the superscript indicates the whole row of the matrix) is equal to $\pi * m^1 + k * m^2$. Using the following matrix format one can solve for π and k: $\begin{bmatrix} 0 & -m_{12} \\ m_{12} & 0 \end{bmatrix} \begin{bmatrix} \pi \\ k \end{bmatrix} = \begin{bmatrix} -m_{13} \\ -m_{23} \end{bmatrix}$.

This gives a value for $\pi = \frac{-m_{23}}{m_{12}}$ and a value for $k = \frac{m_{13}}{m_{12}}$[42].

Once these two are known, then it is clear $m_{34} = (m_{13}m_{24} - m_{14}m_{23}) / m_{12}$ because one can see from visual examination of M that $m_{34} = \pi * m_{14} + k * m_{24}$. Exactly the same kind of logic can be used to show for larger matrices that $m_{ij} = (m_{1i}m_{2j} - m_{1j}m_{2i}) / m_{12}$

[42] There is a slight, innocuous, error in the published paper, where the sign given for k is negative.

for all i,j such that j > i > 2, which is Equation 3-5 in the text and one of the main results of the Browning and Chiappori 1998 paper.

Extending from their results, in the three decision-maker household a 6*6 M matrix will have a rank of 4 because both u and v are no longer vectors, but 6*2 matrices. In the seven good model of Chapter 3 (where again the seventh good is included to preserve adding-up), this 6*6 matrix will look as follows.

$$\begin{bmatrix} 0 & m_{12} & m_{13} & m_{14} & m_{15} & m_{16} \\ -m_{12} & 0 & m_{23} & m_{24} & m_{25} & m_{26} \\ -m_{13} & -m_{23} & 0 & m_{34} & m_{35} & m_{36} \\ -m_{14} & -m_{24} & -m_{34} & 0 & m_{45} & m_{46} \\ -m_{15} & -m_{25} & -m_{35} & -m_{45} & 0 & m_{56} \\ -m_{16} & -m_{26} & -m_{36} & -m_{46} & -m_{56} & 0 \end{bmatrix}$$

Following the same type of arguments as in the 4*4 matrix two-decision maker case, row 5 of the 6*6 matrix is linearly dependent on the first 4 rows: $m^5 = \pi * m^1 + k * m^2 + l * m^3 + q * m^4$. The 4 unknowns π, k, l *and* q can be found,

using the following matrix formulation:

$$\begin{bmatrix} 0 & -m_{12} & -m_{13} & -m_{14} \\ m_{12} & 0 & -m_{23} & -m_{24} \\ m_{13} & m_{23} & 0 & -m_{34} \\ m_{14} & m_{24} & m_{34} & 0 \end{bmatrix} \begin{bmatrix} \pi \\ \kappa \\ l \\ q \end{bmatrix} = \begin{bmatrix} -m_{15} \\ -m_{25} \\ -m_{35} \\ -m_{45} \end{bmatrix}.$$

This is solved by Gaussian reduction, the main steps of which follow:

$$\begin{matrix} 1 & 0 & \dfrac{-m_{23}}{m_{12}} & \dfrac{-m_{24}}{m_{12}} & \dfrac{-m_{25}}{m_{12}} \\ 0 & 1 & \dfrac{m_{13}}{m_{12}} & \dfrac{-m_{14}}{m_{12}} & \dfrac{m_{15}}{m_{12}} \\ m_{13} & m_{23} & 0 & -m_{34} & -m_{35} \\ m_{14} & m_{24} & m_{34} & 0 & -m_{45} \end{matrix}$$

Here, the first two rows of the previous matrix have been swapped and then divided by m_{12} and $-m_{12}$ respectively. Continuing the reduction, one finds:

$$\begin{matrix} 1 & 0 & \dfrac{-m_{23}}{m_{12}} & \dfrac{-m_{24}}{m_{12}} & \dfrac{-m_{25}}{m_{12}} \\ 0 & 1 & \dfrac{m_{13}}{m_{12}} & \dfrac{m_{14}}{m_{12}} & \dfrac{m_{15}}{m_{12}} \\ 0 & 0 & 0 & \dfrac{-m_{23}m_{14}}{m_{12}} + \dfrac{m_{13}m_{24}}{m_{12}} - m_{34} & \dfrac{-m_{23}m_{15}}{m_{12}} + \dfrac{m_{13}m_{25}}{m_{12}} - m_{35} \\ 0 & 0 & \dfrac{-m_{24}m_{13}}{m_{12}} + \dfrac{m_{14}m_{23}}{m_{12}} + m_{34} & 0 & \dfrac{-m_{24}m_{15}}{m_{12}} + \dfrac{m_{14}m_{25}}{m_{12}} - m_{45} \end{matrix}$$

Here, $-m_{13}$ times the first row of the previous matrix has been added to its third row and $-m_{14}$ times the second row has been added to its fourth row. Then $-m_{23}$ times the new second row has been added to the new third row and $-m_{24}$ times the new second row has been added to the new fourth row.

The next matrix leads quickly to the solutions:

$$\begin{matrix} 1 & 0 & \dfrac{-m_{23}}{m_{12}} & \dfrac{-m_{24}}{m_{12}} & \dfrac{-m_{25}}{m_{12}} \\ 0 & 1 & \dfrac{m_{13}}{m_{12}} & \dfrac{m_{14}}{m_{12}} & \dfrac{m_{15}}{m_{12}} \\ 0 & 0 & 1 & 0 & \dfrac{\dfrac{m_{14}m_{25}}{m_{12}} - \dfrac{m_{24}m_{25}}{m_{12}} - m_{45}}{\dfrac{m_{14}m_{23}}{m_{12}} - \dfrac{m_{24}m_{13}}{m_{12}} + m_{34}} \\ 0 & 0 & 0 & 1 & \dfrac{-\dfrac{m_{23}m_{15}}{m_{12}} + \dfrac{m_{13}m_{25}}{m_{12}} - m_{35}}{\dfrac{m_{13}m_{24}}{m_{12}} - \dfrac{m_{23}m_{14}}{m_{12}} - m_{34}} \end{matrix}$$

Here, the third and fourth matrix have been swapped. Then the third has been divided by $\frac{m_{14}m_{23}}{m_{12}} - \frac{m_{24}m_{13}}{m_{12}} + m_{34}$ and the fourth has been divided by $\frac{m_{13}m_{24}}{m_{12}} - \frac{m_{23}m_{14}}{m_{12}} - m_{34}$.

Once this stage is reached, the four solutions are straightforward:

$$\pi = (-m_{25} + m_{23}[F] + m_{24}[G]) / m_{12}\,;\ k = (m_{15} - m_{13}[F] - m_{14}[G]) / m_{12}\,;\ l = F\,;\ q = G.$$

$$\textit{where}\ \ F = \frac{m_{14}m_{25} - m_{24}m_{15} - m_{45}m_{12}}{m_{23}m_{14} - m_{24}m_{13} + m_{34}m_{12}} \quad \textit{and} \quad G = \frac{m_{25}m_{13} - m_{23}m_{15} - m_{35}m_{12}}{m_{24}m_{13} - m_{23}m_{14} - m_{34}m_{12}}.$$

As one can see from the initial 6*6 matrix M,

$m_{56} = \pi * m_{16} + k * m_{26} + l * m_{36} + q * m_{46}$.

This directly leads to the restriction used for the three-decision-maker test in Chapter 3:

$$m_{56}m_{12} = -m_{16}m_{25} + m_{16}m_{23}[F] + m_{16}m_{24}[G] + m_{26}m_{15} - m_{26}m_{13}[F] - m_{26}m_{14}[G] + m_{36}m_{12}[F] + m_{46}m_{12}[G],$$

Chapter 4

Chapter 4: Appendix 1: Tables

Multivariate probit models

Table 4-A 1: Correlations from the multivariate probit models

	Model 1		Model 2
N	238	N	179
Participation of operator, spouse and other (male or female children)		Participation of operator, spouse and male children	
P-values of test for zero correlations		*P-values of test for zero correlations*	
Operator and spouse	0.002	Operator and spouse	0.004
Operator and child	0.212	Operator and child	0.224
Spouse and child	0.001	Spouse and child	0.004

The models contain the same basic set of independent variables as the bivariate probits.
There is a high level of multicollinearity (even in drastically reduced models), and so the correlations above are all that is reported.
Tables are available on request.

Individualised multinomial logit models

Table 4-A 2: Individualised Multinomial Logit Models

Operator	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
Age	0.197 [0.16]	-0.008 [0.85]	0.159*** [0.01]	0.070 [0.26]	0.129** [0.03]
Age^2	-0.002 [0.21]	0.000 [0.79]	-0.002*** [0.01]	-0.001 [0.32]	-0.001** [0.04]
Years of education	-0.040 [0.60]	-0.036 [0.25]	0.141*** [0.00]	0.035 [0.43]	0.051 [0.21]
Years experience on-farm	0.002 [0.85]	-0.019** [0.01]	-0.039*** [0.00]	-0.039*** [0.00]	-0.031*** [0.00]
Average years of education of rest of household	-0.140 [0.12]	-0.027 [0.38]	0.081** [0.03]	0.072 [0.11]	0.076* [0.06]
Average on-farm experience of rest of household	-0.010 [0.63]	0.012 [0.20]	-0.002 [0.90]	-0.005 [0.75]	-0.001 [0.94]
Number of people over 65 in the house	-0.281 [0.72]	-0.697* [0.08]	0.245 [0.43]	0.075 [0.85]	0.500 [0.12]
Number of children (under 15)	0.025 [0.86]	0.028 [0.62]	0.055 [0.43]	0.088 [0.27]	0.067 [0.35]
Number of children (under 7)	dropped				
Unearned income	0.001** [0.02]	0.001 [0.56]	0.001 [0.11]	0.001 [0.83]	-0.001* [0.07]
Number of rooms in the family home	0.173 [0.59]	0.163 [0.14]	0.263** [0.05]	0.293** [0.05]	0.416*** [0.00]
Spanish speaking	-0.199 [0.80]	-0.188 [0.46]	-0.368 [0.22]	-0.283 [0.39]	0.351 [0.28]
Home owners	-2.081** [0.02]	-0.133 [0.67]	-0.299 [0.41]	-0.269 [0.53]	-0.543 [0.16]
Farm capital per adult	-0.001 [0.50]	0.001 [0.18]	-0.001*** [0.00]	-0.001** [0.01]	-0.002*** [0.00]
Farm size per adult	-0.786* [0.08]	-0.047* [0.05]	-0.087** [0.03]	-0.080** [0.04]	-0.031 [0.24]
Change in farm size (1991-1994)	0.53 [0.29]	-0.021 [0.71]	-0.088 [0.35]	0.009 [0.91]	-0.079 [0.29]
Inverted Simpson Index	0.145 [0.63]	-0.052 [0.58]	-0.215* [0.06]	-0.12 [0.37]	-0.474*** [0.00]
Proportion of farm output sold on the market	-7.188*** [0.00]	-0.860** [0.02]	-1.200*** [0.00]	-0.747 [0.13]	-1.445*** [0.00]
Rice	-31.245 [1.00]	-0.624* [0.07]	-1.237** [0.01]	-0.801* [0.07]	-0.883* [0.08]
Maize	-0.785 [0.33]	0.725*** [0.01]	-0.101 [0.76]	-0.524 [0.22]	0.435 [0.24]
Banana	-0.023 [0.97]	-0.238 [0.45]	-0.469 [0.17]	-0.381 [0.34]	0.395 [0.30]
Potato	0.302 [0.83]	-0.597 [0.21]	-0.764 [0.18]	-0.903 [0.17]	-0.699 [0.21]

Operator	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
Importance of local non-agricultural income sources	-1.502	0.175	-0.541	-0.500	1.679**
	[0.61]	[0.78]	[0.55]	[0.66]	[0.02]
Difficulty of access (distance)	0.046	-0.056	0.119	-0.277*	0.366***
	[0.87]	[0.60]	[0.35]	[0.06]	[0.01]
Difficulty of access (time)	-0.193	-0.078*	-0.006	-0.058	-0.006
	[0.17]	[0.06]	[0.88]	[0.32]	[0.90]
Farm tenure (higher means less secure)	0.171	0.027	0.110	-0.029	-0.159*
	[0.32]	[0.68]	[0.15]	[0.76]	[0.09]
Coastal region	-30.471	0.403	-0.886*	-0.469	-1.243**
	[1.00]	[0.25]	[0.06]	[0.37]	[0.02]
Rainforest region	1.117	0.657*	-0.181	0.621	0.193
	[0.34]	[0.05]	[0.67]	[0.16]	[0.67]
Credit	-0.087	0.188	-0.188	-0.42	-0.227
	[0.91]	[0.48]	[0.59]	[0.32]	[0.55]
Unpaid labour exchange is customary	-0.078	-0.253	0.416	0.114	0.417
	[0.92]	[0.33]	[0.20]	[0.76]	[0.23]
Married into Area	-31.361	0.397	-0.366	0.127	0.686
	[1.00]	[0.62]	[0.76]	[0.91]	[0.42]
Intercept	-2.002	1.460	-3.428**	-0.569	-5.361***
	[0.66]	[0.27]	[0.05]	[0.77]	[0.00]
N	886	886	886	886	886

Spouse	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
Age	0.046	0.138	0.236*	0.122	0.193**
	[0.44]	[0.14]	[0.07]	[0.14]	[0.04]
Age^2	-0.001	-0.001	-0.002	-0.001	-0.002*
	[0.25]	[0.16]	[0.11]	[0.31]	[0.09]
Years of education	-0.006	0.053	0.103	0.111*	-0.012
	[0.90]	[0.40]	[0.16]	[0.07]	[0.85]
Years experience on-farm	-0.001	-0.023	-0.066***	-0.097***	-0.098***
	[0.90]	[0.14]	[0.00]	[0.00]	[0.00]
Average years of education of rest of household	0.018	-0.037	0.093	-0.053	0.019
	[0.71]	[0.56]	[0.26]	[0.42]	[0.78]
Average on-farm experience of rest of household	0.003	-0.005	0.042**	0.001	0.002
	[0.76]	[0.77]	[0.02]	[0.97]	[0.92]
Number of people over 65 in the house	0.103	0.286	-0.536	-0.409	0.44
	[0.81]	[0.60]	[0.49]	[0.50]	[0.43]
Number of children (under 15)	-0.114	0.007	-0.126	-0.345*	-0.303
	[0.39]	[0.97]	[0.59]	[0.06]	[0.10]
Number of children (under 7)	0.092	0.062	0.097	0.242	0.21
	[0.64]	[0.81]	[0.78]	[0.37]	[0.43]
Unearned income	-0.001***	-0.001**	0.001	-0.001**	-0.001*
	[0.00]	[0.03]	[0.90]	[0.02]	[0.06]
Number of rooms in the family home	-0.529***	-0.557*	-0.822**	-0.249	-0.24
	[0.01]	[0.06]	[0.03]	[0.35]	[0.40]
Spanish speaking	0.260	0.409	0.673	0.488	0.999**
	[0.47]	[0.40]	[0.27]	[0.31]	[0.05]
Home owners	-0.45	-0.136	-0.674	-0.868	-1.136*
	[0.40]	[0.84]	[0.38]	[0.17]	[0.07]
Farm capital per adult	0.001	-0.001	-0.002**	-0.001	0.000
	[0.95]	[0.19]	[0.03]	[0.13]	[0.67]
Farm size per adult	0.145*	0.05	0.121	0.065	-0.009
	[0.06]	[0.64]	[0.22]	[0.49]	[0.93]
Change in farm size (1991-1994)	0.091	-0.045	0.118	0.092	-0.081
	[0.59]	[0.85]	[0.59]	[0.66]	[0.72]
Inverted Simpson Index	-0.083	0.179	-0.388	-0.110	-0.187
	[0.56]	[0.33]	[0.14]	[0.58]	[0.36]
Proportion of farm output sold on the market	-1.463***	-2.500***	-2.852***	-1.415**	-2.132***
	[0.01]	[0.00]	[0.00]	[0.04]	[0.00]
Rice	0.583	0.327	-0.19	0.539	0.511
	[0.27]	[0.64]	[0.86]	[0.41]	[0.46]
Maize	0.660*	-0.403	-0.547	0.276	1.210**
	[0.10]	[0.51]	[0.49]	[0.62]	[0.03]
Banana	1.175**	0.778	0.855	0.328	1.336**
	[0.01]	[0.19]	[0.21]	[0.60]	[0.05]
Potato	0.047	0.563	0.708	0.071	1.082
	[0.95]	[0.57]	[0.53]	[0.94]	[0.22]
Importance of local non-agricultural income sources	1.133	2.111*	-0.820	0.688	1.916
	[0.31]	[0.09]	[0.70]	[0.64]	[0.14]
Difficulty of access (distance)	-0.291*	-0.567***	-0.003	-0.274	-0.383*
	[0.07]	[0.01]	[0.99]	[0.19]	[0.07]
Difficulty of access (time)	-0.043	-0.06	0.056	-0.108	0.06
	[0.46]	[0.45]	[0.54]	[0.23]	[0.39]
Farm tenure (higher means less secure)	-0.06	0.016	-0.062	-0.072	0.038
	[0.53]	[0.90]	[0.70]	[0.57]	[0.76]
Coastal region	0.589	1.701**	-1.338	-0.589	1.009
	[0.30]	[0.03]	[0.31]	[0.42]	[0.20]

Spouse	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
Rainforest region	-0.407	-0.300	-1.462	-0.766	0.343
	[0.42]	[0.66]	[0.12]	[0.25]	[0.61]
Credit	0.022	0.072	0.453	-0.797	0.334
	[0.96]	[0.89]	[0.49]	[0.20]	[0.52]
Unpaid labour exchange is customary	0.878**	0.185	0.440	-0.274	1.032*
	[0.02]	[0.72]	[0.49]	[0.59]	[0.06]
Married into Area	-1.011	-0.482	1.582*	-0.581	0.594
	[0.17]	[0.63]	[0.08]	[0.52]	[0.46]
Intercept	1.599	-2.251	-2.436	1.255	-3.295
	[0.43]	[0.41]	[0.53]	[0.65]	[0.24]
N	619	619	619	619	619

Kids	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
Age	-0.343***	-0.163*	-0.071	-0.007	-0.062
	[0.00]	[0.06]	[0.39]	[0.96]	[0.57]
Age^2	0.003***	0.001	0.001	0.000	0.001
	[0.00]	[0.25]	[0.51]	[0.86]	[0.53]
Gender (1=male)	-0.855*	0.282	-0.981**	-2.740***	-1.953***
	[0.06]	[0.58]	[0.04]	[0.00]	[0.00]
Attending school	1.099	0.801	0.016	-0.300	0.558
	[0.17]	[0.35]	[0.99]	[0.83]	[0.61]
Years of education	-0.055	-0.096*	0.016	-0.081	0.021
	[0.25]	[0.06]	[0.76]	[0.23]	[0.75]
Years experience on-farm	0.000	-0.070***	-0.134***	-0.225***	-0.117***
	[0.99]	[0.00]	[0.00]	[0.00]	[0.00]
Average years of education of rest of household	0.028	-0.098	0.007	-0.092	-0.183*
	[0.67]	[0.19]	[0.92]	[0.38]	[0.06
Average on-farm experience of rest of household	0.068***	0.019	0.043	0.038	-0.010
	[0.01]	[0.46]	[0.10]	[0.25]	[0.77]
Number of people over 65 in the house	0.988**	1.126**	0.432	0.782	0.107
	[0.02]	[0.01]	[0.34]	[0.18]	[0.86]
Number of children (under 15)	-0.102	-0.031	-0.008	-0.608*	-0.143
	[0.55]	[0.87]	[0.97]	[0.05]	[0.57]
Number of children (under 7)	0.078	-0.05	-0.078	0.357	0.300
	[0.76]	[0.86]	[0.78]	[0.41]	[0.41]
Unearned income	0.001	0.001	0.001	0.001	-0.001
	[0.81]	[0.46]	[0.42]	[0.84]	[0.27]
Number of rooms in the family home	-0.397	0.028	0.113	-0.151	0.800**
	[0.14]	[0.93]	[0.70]	[0.75]	[0.02]
Spanish speaking	-0.768*	-0.221	-0.771	-0.965	-0.582
	[0.09]	[0.67]	[0.13]	[0.21]	[0.43]
Home owners	0.853	0.835	0.415	0.148	2.689**
	[0.19]	[0.25]	[0.57]	[0.88]	[0.03]
Farm capital per adult	0.001	-0.001	-0.001*	0.001	-0.002**
	[0.63]	[0.15]	[0.09]	[0.70]	[0.05]
Farm size per adult	-0.058	-0.106	-0.011	0.029	0.067
	[0.39]	[0.26]	[0.88]	[0.80]	[0.48]
Change in farm size (1991-1994)	0.519**	0.362	0.563**	0.253	0.818***
	[0.01]	[0.11]	[0.02]	[0.50]	[0.01]
Inverted Simpson Index	-0.298*	-0.344*	-0.703***	-0.186	-0.449*
	[0.05]	[0.05]	[0.00]	[0.48]	[0.08]
Proportion of farm output sold on the market	-0.036	-1.368*	-1.132	-0.34	-0.231
	[0.96]	[0.06]	[0.11]	[0.75]	[0.81]
Rice	dropped				
Maize	-0.055	0.134	0.088	0.635	1.249
	[0.92]	[0.82]	[0.88]	[0.42]	[0.12]
Banana	0.021	0.201	0.011	-0.809	1.620*
	[0.97]	[0.75]	[0.99]	[0.39]	[0.05]
Potato	dropped				
Importance of local non-agricultural income sources	3.863*	5.033**	4.287*	4.609*	5.462**
	[0.09]	[0.03]	[0.06]	[0.06]	[0.03]
Difficulty of access (distance)	0.043	-0.052	0.111	0.059	0.618**
	[0.83]	[0.82]	[0.62]	[0.86]	[0.04]
Difficulty of access (time)	-0.099	-0.161*	-0.067	-0.155	-0.001
	[0.19]	[0.07]	[0.42]	[0.25]	[0.99]
Farm tenure (higher means less secure)	-0.093	0.106	0.109	0.179	-0.079
	[0.49]	[0.47]	[0.46]	[0.38]	[0.75]
Coastal region	0.502	0.653	0.002	0.283	-1.69
	[0.41]	[0.33]	[1.00]	[0.77]	[0.13]
Rainforest region	0.284	-0.225	-1.178*	-0.240	-0.576
	[0.64]	[0.74]	[0.09]	[0.80]	[0.53]
Credit	0.231	0.506	-0.164	-1.536	0.214
	[0.67]	[0.38]	[0.78]	[0.22]	[0.78]
Unpaid labour exchange is customary	0.732	0.380	0.856	-0.304	0.735

Kids	Unpaid farm labour	Farm wage	Non-farm wage	Sole trader	Self-employed with paid staff
	[0.16]	[0.51]	[0.14]	[0.72]	[0.32]
Intercept	5.213*	1.551	1.096	1.005	-5.636
	[0.07]	[0.61]	[0.72]	[0.79]	[0.16]
N	842	842	842	842	842

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported. The excluded group is those who farm for returns, and do not work for wages off-farm.

Instrumental variable Models for Table 4.1

Table 4-A 3: Instrumental variable models for table 4.1

	Farm capital	Inverted Simpson Index	Proportion of farm output sold on the market
Land quality proxy	0.003	-0.000**	0.000
	[0.59]	[0.04]	[0.90]
Total no. in family	-144.657	-0.012	-0.023
	[0.15]	[0.88]	[0.26]
Average distance to amenities	-2.126***	0.001**	-0.000**
	[0.00]	[0.04]	[0.04]
Local daily rate for agricultural labour (male)	-108.774*	-0.264***	0.01
	[0.08]	[0.00]	[0.41]
Population of survey 'segmento'	-0.470**	0.000	-0.000**
	[0.02]	[0.10]	[0.01]
No. of Times Public Transport comes to Village	44.334**	-0.005	0.003
	[0.03]	[0.74]	[0.49]
Weekly farm pay (for female)	-2.407	0.100***	-0.039***
	[0.94]	[0.00]	[0.00]
Average household adult age	7.08	-0.015	0.000
	[0.73]	[0.32]	[0.92]
Average household adult age^2	0.036	0.000	0.000
	[0.87]	[0.52]	[0.88]
Average household adult gender	334.357**	-0.011	0.071**
	[0.05]	[0.93]	[0.04]
Years of education	39.254***	-0.012	-0.004
	[0.01]	[0.30]	[0.22]
Years experience on-farm	1.347	-0.001	-0.002**
	[0.74]	[0.83]	[0.01]
Number of adult family members (15-65)	210.002**	0.002	0.017
	[0.04]	[0.98]	[0.44]
Number of children (under 15)	130.702	0.058	0.010
	[0.21]	[0.45]	[0.64]
Spanish speaking	-1.042	-0.197**	0.283***
	[0.99]	[0.01]	[0.00]
Unearned income	0.034	0.001	0.001
	[0.63]	[0.93]	[0.45]
Home owners	137.74	0.048	-0.073**
	[0.34]	[0.65]	[0.01]
Number of rooms in the family home	119.231***	0.01	0.014**
	[0.00]	[0.69]	[0.05]
Farm size	43.474***	0.014***	0.000
	[0.00]	[0.00]	[0.81]
Change in farm size (1991-1994)	-57.203**	-0.012	-0.012**
	[0.01]	[0.49]	[0.01]
Coastal region	53.494	-0.513***	0.380***
	[0.73]	[0.00]	[0.00]
Rainforest region	-390.772**	-0.315**	0.220***
	[0.02]	[0.01]	[0.00]
Importance of local non-agricultural income sources	1,501.700***	-0.661***	-0.138**
	[0.00]	[0.00]	[0.01]
Difficulty of access (distance)	66.354	0.058	0.011
	[0.20]	[0.13]	[0.31]
Difficulty of access (time)	23.334	0.051***	0.007*
	[0.21]	[0.00]	[0.07]
Farm tenure (higher means less secure)	-40.712	-0.008	-0.018***
	[0.18]	[0.72]	[0.01]
Maize	96.202	-0.206**	-0.019
	[0.44]	[0.03]	[0.45]
Potato	118.257	-0.246**	0.000
	[0.38]	[0.01]	[1.00]
Rice	78.42	-0.381***	-0.017
	[0.62]	[0.00]	[0.61]
Banana	-505.250**	-0.171	-0.220***

	[0.02]	[0.27]	[0.00]
Intercept	-1,966.698***	4.225***	0.605***
	[0.00]	[0.00]	[0.00]
N	946	946	946
R^2	0.177	0.21	0.444

*=significant at 10%, ** at 5% and *** at 1%. P-values in square brackets. Raw coefficients reported.

Chapter 4; Appendix 2. Variables used in the reduced form participation and hours models (with reference also to the models of the other chapters)

The independent variables in each model of chapter 4 (and in other chapters) are measured at up to five different levels – at 'segmento' level (there are normally around 12 households sampled from each segmento, or community), at farm level, at the level of the household, at the level of 'other' members of the household and at individual level. The independent variables are reasonably standard in the literature, although some have rarely been tested for. The independent variables at the different levels are:

Segmento or Community Level (answered by community leaders):[43]

a) Importance of agriculture to the community (a subjective ranking by a committee in each village, ranging from 1=very important to 4=not important).

b) Quality of road access to the area (1=good; 4=bad),

c) Distance in kilometres to a group of amenities – post, phone, market etc.,

d) Distance in time to the nearest public transport,

e) The amount of dwellings in the segmento,

f) Local prices,

g) A 'minka' dummy.

Variables b), c), and e) were dropped in most of the results reported here, due to some multicollinearity leading to difficulties in disentangling effects.

Local prices are given for a range of consumer goods at community level (not for all segmentos, and not for all goods). These prices (for six goods) are used in the AIDS models of Chapter 3. The price of the composite reference 7th good is the average community price for 15 other food goods consumed in all the tested communities. The data allows estimation for 69 out of 111 communities.

[43] In some cases there are two sets of answers for a single segmento code (two communities are in the segmento); for these, simple averages are taken.

The 'minka' dummy is equal to 1 if 'informal' unpaid exchange of agricultural labour takes place in the community. It is equal to zero otherwise.

Farm Level:

a) Main crops produced (4 separate dummies – main crop 1 (maize), main crop 2 (papa – potato), main crop 3 (rice) and main crop 4 (bananas),

b) Farm size,

c) Change in size since 1991,

d) Value of farm capital,

e) Farm tenure (1=secure, ranging to 6=insecure),

f) Proportion of crop output sold,

g) The Inverted Simpson Index of Crop Diversification[44], used as a measure of diversification (see Malchow-Møller and Svarer, 2002),

h) Crop specific farm inputs (used in the models of Chapter 6),

i) The amount of hired labour.

For variables a), f) and g) self reported individual crop prices are used, so the 'main crop' variable, for instance, is not the crop with the largest hectarage on a particular farm, but the most valuable crop. For variables d) (capital) and h) (crop specific inputs) national median prices are used to attempt to get closer to physical quantities. Capital is measured (following Jacoby, 1993) as half the value of traction livestock, plus the value of physical implements. The quantity of hired labour is the reported cost of hired labour divided by the local male agricultural wage, given in the community questionnaire.

Farm profits and livestock outputs and inputs are not included in the models here. Data on some livestock output is available, and on livestock inputs, much output data has to be imputed. The more reduced model form estimated here includes only characteristics of the farm and of farm output, through which farm profits are

---

[44] Diversification Index = Inverted Simpson Index = $\frac{q_i^2}{\sum_{s=1}^{n_i} q_{is}^2} \leq n_i$, where n_i is the number of different crops harvested by the household i in 1994, q_{is} is the value of crop s and $q_i = \sum_{s=1}^{n_i} q_{is}$.

generated. It was decided that the complications and uncertainties involved in specifying a particular functional form for production and possible data inadequacies in monetary values made it more sensible to concentrate on the non-pecuniary variables that are actually chosen.

House and Household Level:

a) House Tenure Dummy (1=Strong Tenure)

b) Number of Rooms in the House

c) Unearned Income (unexpected – inheritances, gifts etc.)

d) Unearned Income (expected – regular transfers, rent, remittances)

e) Savings (plus money owed to the household)

f) The Main Language Spoken in the Household (1=Spanish; 0 = Other)

g) The Number of Children (under 15 and under 7) in the house

h) The Number of Elderly in the House (over 65)

i) Household Credit Dummy (1=In Receipt of Credit)

j) The number of adults in the house

k) Total household expenditure.

'Unearned income' is a key variable theoretically in the farm-household model, and in household models generally, because it provides the link between one-period and multi-period models (Blundell and Walker, 1986). However, in a pared-down reduced form model, like most the models estimated here, it was preferred not to deal with the possible endogeneity involved if 'expected' unearned income is included on the right hand side even if Schultz, 1990, argues that, despite many untested claims to the contrary, the empirical evidence is strongly on the side of unearned income being exogenous.

An alternative to using expected unearned income itself is to use household expenditure as an indicator of permanent income (Jacoby, 1995). This variable was experimented with in the household level models (reported below in Table 4-1). It was found to have a relatively high correlation with the 'number of rooms' variable (a correlation coefficient of 0.4), and to mimic its coefficient pattern in regressions, and it was also correlated quite strongly with 'expected' unearned income (.35), and had insignificant correlations with other variables such as number of adults, total family size, and in particular unexpected unearned income (the correlation coefficient was less than 0.05 and insignificant). The suggestion is that all three high correlation variables, as one might expect, are closely related, since asset position (proxied by number of rooms) is closely related to unearned income which, in turn, is closely related to permanent income (proxied by total household expenditure), with earned income controlled for. It was decided to proceed with the asset position variable(s), which are less likely to suffer severe measurement error and which, perhaps, are slightly less likely to be endogenous. It was thought too that useful information regarding capital market efficiency could be deduced from the significance or otherwise of the house size variable in some of the estimations. Variables d), e) and i), therefore, were dropped from the models (although variable i) was retained in some exploratory regressions, reported below in footnote 23). Variables a), b), c) and j) were retained. The total number of adults (over 14 years old) in the house is also included as a control variable in household and couple level models.

Total household expenditure (which includes the value of auto-consumption) is actually the dependent variable in the permanent income models of Chapter 6.

Individual Level:

a) Education – dummies for completion of 3 years schooling, 8 years, 13 years and 3^{rd} level education, or – more often – education measured in years

b) Experience of On-Farm Work, measured in years

c) Age and Age Squared

d) Sex (1=Male)

e) Dummy for whether a Main Member of the Household or not (Head or Spouse); or Dummies for Main Male, Main Female, other Males (95% of whom are under 40) and other Females (90% of whom are under 40). The other male and female categories are jointly referred to below as adult children, as the vast majority are children of the main couple.

f) At School – a dummy = 1 if a 15-18 year old is a full time student

g) Hourly wages; calculated as earnings divided by hours of wage work (use din the models of Chapter 5)

h) Hourly returns to self-employment; calculated as returns divided by hours in self-employed work (the figures, as seen in Chapter 2, Table 2-11, seem fairly unreliable)

i) Hours in various kind of work – on-farm, off-farm wage work, off-farm self – employed work.

Other Variables: The individual level education and experience variables are averaged (for all relevant adults) for the household level models, while averages for 'others' in the house are included in model c). These variables are important in that they can cast some light on intra-household allocation of labour. If such allocation is carried out according to comparative advantage within the household, we would expect the coefficients of these variables to be opposite in sign to those on the same variables at individual level.

The education level of the household head and of the most educated person in the house are also included in some versions of the model. Age and sex 'averages' are also included in the household model. Per adult averages of three quantitative variables at farm and household level (farm size and capital, number of rooms in the house) are included in model c). There are also regional dummies for coastal and jungle regions, although the main-crop dummies come fairly close to being regional identifiers (nearly all banana farms are in the rainforest region; nearly all rice farms in the coastal region; the vast majority of maize and potato farms are in the sierra). The highlands region (sierra) is generally the excluded category in the reported estimations.

A number if variables are used as instrumental variables in Chapters 4, 5 and 6. These include some of the unused community level variables above, whether a person was born in a city or not, whether they married in or not, the farm size in 1991, and some family personal endowment variables – number of men, number of women etc. They are referred to when appropriate.

Intra-Family 'Bargaining' Variables: Two further variables (a dummy for an operator or spouse marrying into the area and the segmento ratio of men to women) are also included in some estimations to test for the possibility that the unitary household utility framework is inappropriate. Conclusions cannot be drawn on this question on the basis of reduced form estimations such as these, since the two variables in question may have other 'meanings' than the assumed ones – people marrying into an area may, for example, be more outgoing and entrepreneurial than others, thus negating any possible weakness in their bargaining positions over intra-familial allocations. Nevertheless, the availability of these variables in the dataset meant that they could be included in the estimations, and so alternative estimations are run for most models of Chapter 4 with these variables included.

Endogeneity: Some potential variables of interest not included for estimation have already been mentioned (farm profits, expected household unearned income). Others include: returns (or potential returns) to off– (and on-) farm work, the hours actually worked off-farm by others in the household, on-farm hours worked by self and others, hired-in on-farm hours and household consumption. Some (or all) of these variables should be included in more structural models (Pascal et al, 2002; Kimhi and Lee, 1996; Lundberg, 1992; Huffman, 1988), and appear in other chapters, but none of these endogenous variables are appropriate for the reduced form comparisons carried out below.

Almost all the variables actually included might, of course, also be legitimately regarded as being at least partly endogenous to the participation decision. Ahitov and Kimhi (2002), for example, have modelled farm capital as endogenous, using panel data. The risk management variable (the inverted Simpson index) and the cash-flow variable (the proportion of crops sold on the market) have been modelled as being exogenous by Malchow-Møller and Svarer (2002) and are also modelled as exogenous here, but both variables may be endogenous. They could potentially be instrumented by the 'main crop' dummy variable (expected to have a direct influence itself on off-farm choices, and therefore included in the model) and by some unincluded segmento-level variables (in particular, the price of male and female farm labour and the amount of dwellings in the segmento, as well as distances from various services). However, the 'uninstrumented coefficient' on the two variables is of at least as much interest as their possible truly exogenous effect (if any), and it has been decided not to instrument them in any of the reported estimations except for illustrative reasons in alternative versions of the first two household level estimations

(Tables 4.1 below), and to trust that any bias incurred is not too high a price to pay for the probable gain in information.

The main wealth indicator variable (the number of rooms) could also be endogenous, being a possible result of past off-farm efforts. If it is, a positive sign in participation estimations is expected. If it is seen as exogenous, however, a negative sign is expected. Furthermore, to the extent that capital markets are imperfect we would expect the wealth indicator to influence entry into self employment more than, or rather than, entry into better paying wage employment, given that extra rooms may provide both space and collateral for self-employed work, apart from directly indicating the availability of cash. Attention, consequently, is fixed on its possibly differentiated effects on participation in wage and self employment in the later, more disaggregated models.

This variable is not instrumented in any of the IV estimations below because unexpected unearned income can be fairly relied upon to be exogenous, and because convincing instruments cannot be found.

Dependent Variables: For the household level model (model a)), the values of the dichotomous variable represent household participation in off-farm work (one) and on-farm work only (zero). The cut-off point for a household to be assigned a value of one is that total household off-farm hours are greater than one hundred per year.

For the operator/spouse model (model b)), the value of one is assigned to either individual if they carry out any off-farm hours whatsoever. Only households with both a (male) operator and (female) spouse are included in this model, giving a sample size of five hundred and thirty two couples. Sample sizes are smaller still when the model is later extended to include working adult children (Table A1, in the appendix).

The values of the dependent variable in the dichotomous models individual models (model c)) represent off-farm paid and on-farm *paid* or *unpaid* work. Those of the dependent variable in the trichotomous models represent: *on-farm paid* work – at least 95% of personal income comes from on-farm work – *on-farm unpaid* work (no money is received for working on the family farm, and no off-farm paid work is done) and *off-farm paid* work. Finally, in estimations where there are six categories of the dependent variable, the category into which a person is placed is based on the following: *on-farm paid* – at least 95% of personal income comes from on-farm work;

on-farm unpaid – main unpaid work is on the farm: does no paid work anywhere; *farm wage labour* – those not in the first two categories whose single greatest source of off-farm income is wages from agricultural work; *non-agricultural wage labour* – those not in the first categories whose single greatest source of off-farm income is wages from non-agricultural work; *sole trader* – those not in the first categories whose single greatest source of off-farm income is earnings from small-scale self-employment; *self-employed employer* – those not in the first categories whose single greatest source of off-farm income is earnings from larger-scale self-employment (i.e. with at least one paid employee).

Finally, the software used in the analyses are: Stata 8 for most of the estimations in chapters 3, 4 and 5, and the permanent income regressions of chapter 6; Limdep for the multivariate probit estimation in chapter 4, as well as the bivariate selection models in the next chapter 4, appendix 3; SPSS for most of the summaries in chapter 2; GAMS for the entropy estimations in chapter 3 and the DEA analysis in chapter 6.

Chapter 4; Appendix 3. Reduced form hours estimations

Introduction

One reason not to estimate hours equations is that, like income variables, the quality of the data on 'time at work' variables is notoriously unreliable in household surveys. Even with reliable data, it is arguable whether reduced form hours equations would add substantially to the information already garnered, and the normal procedure in modelling farm household labour supply models is to pass directly from one or other of the reduced form participation estimations of the previous sub-section to earnings equations, and from these to more 'structural' hours equations (which will usually include fitted returns to off-farm work, estimated from the earnings equations, on the right hand side). However, there is one respect, at least, where reduced form equations of off-farm hours might be a useful fall-back or guide to the plausibility of more structural hours equations: the data on hourly or daily returns on which some structural equations are based is likely to be even more problematic than the 'time' data used to describe the dependent variable. Also, given that, as in this study, the unpaid category always has no reported returns at all, and that reported returns for farm work may be especially unreliable (so structural models including even just the second of these categories will have to be problematic) pared down reduced form hours equations may be all that can be estimated whenever full structural models are simply impossible, or judged to be highly unreliable.

Also, the usual three step process – participation equation – wage equation – hours equation – is not normally taken account of in the calculation of standard errors for the structural hours equation. This calculation tends to be based on a two stage procedure (participation and wages, or participation and hours), or even a one stage procedure, adding to potential unreliability in the interpretation of more 'structural' models.

Labour market considerations and interpretation of reduced form coefficients

Apart from the specific information on the variables' coefficients that one gains from a reduced form hours estimation, some indirect light may also be cast on the workings of the labour market. In reduced form individual level hours models such as those estimated below, if a variable has a positive sign it cannot be determined if this occurs because a) the variable leads to higher hourly rewards, and labour supply is forward bending, b) the variable leads to lower hourly rewards, and labour supply is backward bending, c) the variable, which may or may not also determine rewards, increases off-farm labour supply at given rewards (e.g. more education might increase hourly rewards; but more education may also separately affect the disutility of work, perhaps, or the ability to negotiate overtime, which in turn might also affect hours worked, independently of the effect through rewards – thus affecting both labour demand and labour supply). In many cases, it is clear what effect we expect a variable to have on rewards – being very young is likely to reduce them; having more education is likely to increase them. We may also surmise variables' effects on the desire and ability to work longer hours off-farm for given rewards: more education may increase these; being near public transport may also increase them; having young children may decrease or increase them – depending on whether one is physically caring for the children or not; factors that increase farm profit, such as having a having a large farm or high level of farm capital, should decrease them; being in certain regions, or being constricted by altitude or environment to grow certain crops may also increase/decrease them – we can speculate, for instance, that growing rice will decrease them.

More generally, it is possible to summarize some reduced form coefficient signs with regard to what they might mean for labour supply. A negative sign, for example, in pared down reduced from hours models for variables that we expect to have a positive effect on rewards to work, and a neutral or positive effect on labour supply for given rewards, is indicative – if not contradicted by other variables – of backward bending labour supply. A positive sign for variables that we expect to have a negative effect on rewards, and a neutral or negative effect on labour supply for given rewards, may also be indicative of backward bending labour supply. A forward bending supply curve may be indicated if the coefficient is positive and the variable is likely to increase rewards and have a neutral or positive effect on labour supply for a given reward. Finally, a forward bending supply curve may also be indicated if the coefficient is negative and the variable is likely to decrease rewards and have a neutral or negative effect on labour supply for a given reward.

If different variables that affect both returns and labour supplied for given returns lead to contradictory conclusions concerning the shape of the labour supply curve, then nothing more can be said without incorporating returns variables into the estimations, and moving towards a more structural model. This is also true if we expect a variable to affect rewards in a different direction from that of its effect on the amount of labour supplied for given rewards. The direction of the labour supply curve cannot, therefore, generally be inferred from pared down reduced form models. However, if a clear majority of variables fit into the first two or the second two of the (non-exhaustive) four cases above, then an educated guess on the slope of the supply curve may be hazarded.

Switching of signs between participation and hours models (positive in participation and negative in hours, or negative in participation and positive in hours) could indicate some form of rationing within the family or in the labour market (e.g. if education has a negative coefficient in highly paid non-agricultural wage labour, and a positive coefficient in the hours estimation), although rationing may also exist if signs are the same in both estimations (e.g. if the hours coefficient were negative in the last example, rationing of entry could co-exist with a backward bending labour supply curve): in this case, as in others, judicious comparison of both hours and participation estimations may help us decide if there are labour market problems (e.g. two possible interpretations of the last example are rationing plus backward bending labour supply versus a simple negative effect on rewards: in any one situation, one of these is likely to be more plausible than the other). Switching of signs is also, of course, consistent with a smooth functioning of the labour market (e.g. a characteristic may make it likely for someone to enter a particular category, but may also reduce rewards relative to others), so each case needs to be viewed in context: nevertheless, it may often be useful to consider the signs of both estimations together when reading hours estimation results.

Finally, if one finds significant negative values for the inverse Mills ratio in a sample selection model below, the intuition is that unknown factors that encourage (discourage) participation also discourage (encourage) hours worked. Some missing variable or variables have opposite effects: the logic is the same as for known variables discussed in the last paragraphs – and, again, the possible reasons include such factors as backward bending labour supply, some form of intra-familial rationing or, again, it could also be because of rationing or imperfections in the labour market, or other markets that impinge on it.

Hours estimations

The same set of X variables is used in the hours estimations as were used in the participation estimations. Since each of the three main sets of equations (at household, couple and individual family member level) use Heckman's sample selection procedure (augmented in the multinomial logits by procedures attributable to Bourguignon et al, 2001)[45], the problem of correctly specifying the selection procedure is common to all. One solution in the literature to this problem is simply to use the non-linearity of the initial probits / multinomial logits as the criterion guaranteeing solution of the joint likelihood function. However, a more intuitively appealing specification, because less obviously arbitrary, is to find variables that can plausibly be excluded from the hours equation, but not from the participation equation. Given that we have included work in self-employment as an activity in the non-agricultural sector, it is particularly difficult to find such identifying variables. Regional dummies, and two of the 'main crop' dummies (excluding rice and bananas) were used as identifying variables, and are excluded from the hours estimations. It must be conceded that this choice was made after much experimentation, and not on a priori grounds. However, it was the only choice which permitted a sensible value of 'rho' in the two-stage Heckman procedure[46].

The hours equations have been estimated using two-step sample selection models for households as a whole (Table 4.A.4), operators and spouses together (Table 4-A.5) and individuals within the household (Table 4.A.6). The sample, selection variables (lambda's) are significant positive (at 10%) in the household model, and in a number of cases in the individual model.

---

[45] LR tests of the household level model indicated that Tobits were inappropriate.
[46] Rho was 1, or very close to 1, for all other specifications.

Table 4-A 4: Household Selection Model

Dependent variable	Heckman Selection Model Household Hours
Average age of adults	85.580** [0.02]
Average age^2 of adults	-0.908** [0.03]
Average sex of adults	317.764 [0.43]
Average education of adults	76.853*** [0.00]
Average experience on-farm of adults	-41.174*** [0.00]
Number of adults in family	505.772*** [0.00]
Number of family members in the home	59.318 [0.15]
Number of adults not in the home	-92.496 [0.65]
Spanish speaking	324.961** [0.04]
Unearned income	-0.230** [0.01]
Home owners	-132.967 [0.52]
Farm capital	-0.311*** [0.00]
Farm size	-23.578** [0.02]
Change in farm size (1991-1994)	57.652** [0.01]
Proportion of farm output sold on the market	-456.511** [0.01]
Inverted Simpson Index	-245.653*** [0.00]
Importance of local non-agricultural income sources	621.015* [0.09]
Farm tenure (higher means less secure)	7.361 [0.87]
Coastal region	-170.235 [0.46]
Rainforest region	150.041 [0.52]
Rice	12.47 [0.96]
Maize	69.343 [0.71]
Banana	-554.277** [0.03]
Potato	485.092 [0.14]
Intercept	-2,413.723** [0.01]
N	1002
Wald	297.29
Test for lambda	0.000
p-value	0.00

When we examine Table 4-A.4 we find that in no case is there any full sign reversal between the participation estimations of Table 4-1 and hours equations. The factors that cause households as a whole to do some work off-farm are also, it appears, the factors that encourage them to work longer hours off-farm. Factors such as unearned income and the farm level variables have clear, and quite strong, effects. Education has an insignificant effect. In unreported estimations, average household education is replaced by education of the household head, and again, by education of the most educated household member. In the first of these cases education is found not to be significant, while in the second it is significant (positive). This slight leaning towards a positive coefficient appears to be contradicted by some of the later results (in Table 4-11) and the complications found by Lazlo (2002) concerning education's allocational and intra-household spillover effects are perhaps related to these contradictions. Farm size has the expected negative impact on hours.

Table 4-A 5: Hours equations based on the bivariate probit selection process

Dependent variable	Operator (spouse working off farm)	Operator (spouse not working off farm)	Spouse (operator not working off farm)	Spouse (operator working off farm)
Age (operator)	66.418 [0.26]	-23.793 [0.72]	-116.295 [0.66]	-29.531 [0.65]
Age^2 (operator)	-0.758 [0.22]	0.313 [0.66]	0.895 [0.69]	0.357 [0.60]
Age (spouse)	-4.403 [0.94]	96.317 [0.11]	-65.795 [0.71]	-23.545 [0.71]
Age^2 (spouse)	0.329 [0.61]	-1.094 [0.14]	0.755 [0.74]	0.435 [0.56]
Years of education (operator)	-25.207 [0.29]	32.414 [0.17]	29.346 [0.78]	-3.78 [0.89]
Years of education (spouse)	39.323 [0.12]	20.232 [0.42]	-98.535 [0.25]	-16.421 [0.53]
Years experience on-farm	-10.567 [0.36]	1.706 [0.92]	-98.715 [0.23]	-3.42 [0.78]
Years experience on-farm (spouse)	-8.242 [0.65]	-19.399 [0.53]	155.229 [0.27]	-25.959 [0.19]
Number of adults	-148.200 [0.03]	-85.816 [0.17]	308.076 [0.49]	-67.958 [0.39]
Family size	235.773*** [0.00]	78.678 [0.25]	-505.335 [0.24]	142.135* [0.09]
Unearned income	-0.003 [0.96]	0.215 [0.11]	-0.625 [0.14]	-0.049 [0.47]
Farm capital	-0.292** [0.06]	-0.059 [0.53]	-0.015 [0.96]	-0.162 [0.18]
Farm size	-21.086* [0.08]	5.759 [0.61]	-39.75 [0.26]	-11.235 [0.37]
Change in farm size (1991-1994)	18.597 [0.54]	-0.072 [0.95]	-1.126 [0.52]	-0.403 [0.73]
Inverted Simpson Index	-169.706* [0.05]	-108.051 [0.16]	-80.568 [0.74]	27.586 [0.78]
Proportion of farm output sold on the market	121.282 [0.62]	-289.112 [0.15]	132.991 [0.90]	-10.447 [0.96]
Spanish speaking	219.66 [0.16]	362.256** [0.04]	96.832 [0.90]	16.345 [0.93]
Home owner	-227.166 [0.19]	18.178 [0.94]	1647.903 [0.19]	-100.317 [0.60]
Importance of local non-agricultural income sources	237.572 [0.49]	-15.443 [0.97]	2702.919 [0.31]	50.73 [0.89]
Farm tenure (higher means less secure)	-69.823* [0.07]	2.468 [0.95]	183.899 [0.19]	-16.738 [0.70]
Coastal region	-16.521 [0.96]	62.274 [0.83]	-210.148 [0.81]	-330.465 [0.26]
Rainforest region	-135.443 [0.57]	-473.641* [0.10]	1455.82 [0.30]	349.645 [0.21]
Rice	-194.02 [0.48]	-220.467 [0.38]	715.661 [0.43]	-130.366 [0.67]
Maize	-104.584 [0.66]	-312.299 [0.26]	1448.754 [0.33]	160.436 [0.56]
Banana	242.649 [0.43]	-151.467 [0.53]	-638.69 [0.42]	67.826 [0.84]
Potato	126.659 [0.68]	-332.884 [0.44]	-369.818 [0.75]	-362.323 [0.20]
Intercept	-857.199 [0.57]	-434.909 [0.79]	4458.101 [0.61]	1877.399 [0.22]
Lambda for operator	19.551 [0.98]	-533.177 [0.47]	-4676.1 [0.26]	126.419 [0.78]
Lambda for spouse	451.137 [0.39]	534.92 [0.47]	4684.7 [0.26]	-123.999 [0.79]
N	162	194	45	164
Adjusted R^2	.1	.15	.25	.13

Comparing the hours model of Table 4-A.5 with the bivariate probit participation model of Table 4-2 we find, again, that there are no sign reversals between the hours and participation results. The additional information that we do gain from the hours equation is that having children appears to lead to more off-farm work for both operator and spouse, once they are working off-farm. On the other hand, if there are more adults in the family, the (male) operator is likely to reduce his hours off-farm. The farm size and capital variables both show through strongly for operator and spouse – again suggesting some simultaneity in developing on and off farm strategies and possibly, to the extent that capital and farm size are exogenous, some use of increased hours of off-farm work by both operator and spouse to make up far having smaller, less capitalised farms. Once again, as we have already seen, risk management using on-farm and off-farm choices does not appear to be undertaken by the spouse. More diversified operators work less hours off the farm. Finally, as already noted, the sample selection variables (Huffman and Lange, 1988[47]) are not significant in any of the four estimations.

Table 4-A 6: Off-farm hours (in 4 categories)

Dependent variable: off-farm hours.......in.......	Farm wage work	Non-farm wage work	Sole-trading self-employment	Self-employment as an employer
Age	3.736 [0.87]	25.336 [0.49]	-36.186 [0.38]	23.837 [0.60]
Age^2	0.051 [0.84]	-0.369 [0.38]	0.245 [0.59]	-0.403 [0.39]
Attending school	-761.160*** [0.00]	-755.269** [0.02]	-1,135.929* [0.09]	780.988 [0.19]
Spouse	-249.425 [0.50]	-12.924 [0.98]	-1,042.52 [0.13]	94.801 [0.90]
Other male	198.945 [0.60]	-350.395 [0.41]	-386.493 [0.68]	281.651 [0.64]
Other female	-205.639 [0.63]	-244.625 [0.54]	-1,216.060* [0.08]	28.238 [0.96]
Years of education	-33.537* [0.08]	1.147 [0.97]	-51.473 [0.24]	-71.469** [0.05]
Years experience on-farm	-10.440 [0.16]	-0.967 [0.93]	-31.309 [0.13]	-16.658 [0.20]
Average years of education of rest of household	0.503 [0.94]	10.323 [0.24]	17.672 [0.16]	1.145 [0.92]
Average on-farm experience of rest of household	-5.639 [0.22]	2.277 [0.71]	3.522 [0.66]	1.760 [0.81]
Number of people over 65 in the house	-209.382* [0.06]	-479.747*** [0.01]	-59.397 [0.82]	-48.74 [0.80]
Number of children (under 15)	368.702* [0.09]	-365.792 [0.32]	135.118 [0.73]	180.673 [0.65]
Number of children (under 7)	238.567*** [0.00]	239.067** [0.04]	49.382 [0.74]	71.588 [0.58]
Unearned income	-0.083 [0.46]	0.154 [0.39]	0.096 [0.76]	-0.257 [0.25]
Number of rooms in the family home	-53.448 [0.38]	-41.622 [0.63]	-12.343 [0.93]	-20.864 [0.86]
Spanish speaking	299.603*** [0.01]	371.009** [0.02]	123.702 [0.59]	200.515 [0.38]
Home owners	130.685 [0.34]	143.766 [0.45]	489.412* [0.06]	132.973 [0.62]
Farm capital per adult	-0.107 [0.28]	0.089 [0.66]	-0.192 [0.24]	-0.307 [0.35]

[47] Various combinations of bivariate selection variables were tried – and all were found insignificant.

Dependent variable: off-farm hours.......in.......	Farm wage work	Non-farm wage work	Sole-trading self-employment	Self-employment as an employer
Farm size per adult	3.599 [0.82]	-35.128 [0.27]	52.760* [0.09]	-49.692** [0.04]
Change in farm size (1991-1994)	61.741 [0.13]	72.948 [0.25]	-126.973 [0.11]	102.006** [0.03]
Inverted Simpson Index	-143.882*** [0.01]	11.44 [0.88]	-161.889 [0.19]	-16.111 [0.88]
Proportion of farm output sold on the market	274.372* [0.08]	411.217* [0.10]	-219.914 [0.53]	-134.36 [0.68]
Rice	192.46 [0.30]	564.544** [0.05]	-538.266 [0.12]	-316.073 [0.31]
Banana	-350.046 [-0.10]	-206.001 [0.52]	-697.064* [0.08]	110.113 [0.74]
Importance of local non-agricultural income sources	-154.424 [0.53]	-657.162 [0.11]	809.198 [0.19]	65.94 [0.90]
Difficulty of access (distance)	9.417 [0.60]	-111.837 [0.15]	34.718 [0.81]	-189.624 [0.11]
Difficulty of access (time)	39.963 [0.14]	9.974 [0.72]	-48.404 [0.42]	-33.902 [0.34]
Farm tenure (higher means less secure)	-71.476** [0.01]	-25.29 [0.56]	111.280* [0.06]	17.186 [0.75]
No. of men relative to women in area-men	4.081 [0.73]	-2.265 [0.91]	4.872 [0.81]	23.298 [0.33]
No. of men relative to women in area - women	-11.354 [0.77]	-685.129 [0.26]	29.932 [0.59]	49.767 [0.93]
Migrated into area to get married	459.64 [0.23]	1,376.954** [0.02]	-434.29 [0.39]	-454.642 [0.36]
Selection variable 1	1,899.576** [0.04]	2,249.755* [0.09]	-1,281.34 [0.52]	-343.149 [0.84]
Selection variable 2	1,086.82 [0.31]	315.142 [0.76]	-1,372.69 [0.47]	2,184.88 [0.13]
Selection variable 3	195.727 [0.54]	-384.669 [0.73]	2,413.80 [0.31]	4,065.444** [0.01]
Selection variable 4	-356.104 [0.72]	-95.029 [0.81]	385.009 [0.84]	1,287.03 [0.46]
Selection variable 5	-101.595 [0.94]	2,173.38 [0.23]	-417.038 [0.68]	2,431.76 [0.30]
Selection variable 6	2,200.024** [0.05]	-1,293.16 [0.39]	-275.606 [0.90]	27.89 [0.96]
Intercept	2,104.393** [0.01]	2,269.51 [0.13]	2,827.63 [0.32]	4,342.773** [0.04]
N	397	306	174	221
R^2	0.28	0.231	0.292	0.176

When the multinomial based selection model of Table 4-A.6 (Bourguignon et al, 2001) is compared with the multinomial logits in Tables 4-7[48], we find that there is some switching in signs between participation and hours equations. The 'at school' variable tends to place individuals in the 'hours in unpaid farm-work' category, but – not surprisingly – to reduce their hours once they are in it. Rice farmers who work off-farm tend to work longer hours. The farm size variable has a negative effect on entry to the 'sole trader self-employed' category (category five), but a positive effect on hours worked once in the category. This is not easy to explain in a reduced form model, as a negative sign would be expected for the hours worked variable (larger farm size should increase farm profits, and reduce the supply of off-farm work), but

[48] An hours regression was also run for the seventh category – those specialising in home production – and being at school reduces hours, being the spouse increases them enormously (about 30 hours per week extra), while being from more diversified farms in less agricultural areas appears to reduce them – suggesting some trade-off or blurring of the borders between farm and house work. The selection variable for category seven is significant negative (those who 'surprisingly' enter the category do less hours than expected for people with their characteristics).

maybe this is outweighed by reduction in cost factors (e.g. credit or manufacturing costs might be cheaper on large farms).

The farm growth variable has a positive effect in category six (self-employed with paid staff). Apart from this estimation (and that of Table 4-A.2), this variable has been completely insignificant. The result in Tables 4-A.6 suggests that certain individuals from households that have been expanding their farm also work longer hours off-farm. This is the first suggestion in all the estimations that off-farm and on-farm work may be complementary, over time, for some people. It could be that increasing farm size is driving this process, but the fact that significance arises only in hours rather than participation estimations suggests the reverse: certain people appear to be using their rewards from off-farm work to increase the size of their farms (rather than using their recently expanded farms to get capital for off-farm work).

The effect of having children is also interesting in these estimations: more children reduce adult unpaid hours on the farm, suggesting that these children take up the slack themselves; more young children (especially those under seven years of age) mean longer hours in wage work, but they are insignificant with regard to self-employment (where, again, they may help their parents and where, as seen in Table 4-7, they have contradictory effects on the entry into small scale self-employment of operators and spouses).

Having old people in the house appears to have negative effects on wage work. Time may be needed for carers to be at home, or the old people in the house may contribute some unpaid work themselves, allowing an increase in leisure for other adults.

Secure household tenure has positive effects on hours in the sole trader category. As we have seen, this job category is relatively important for women; tenure may improve access to credit.

Education generally has negative effects on self-employed hours, suggesting the possibility of backward bending labour supply curves in self-employment (and contradicting, as noted, the results in Table 4-9). Speaking Spanish has a positive effect on hours in paid non-agricultural employment. Selection variables are significant in categories three, four and six, and the method of Bourguignon et al (which allows for cross-category selection effects) is justified, since the significant categories tend not to be the category in which the person is found. The patterns, however, are difficult to interpret.

The second part (the hours equations) of the selection model have been tested for individual effects, using the Breusch-Pagan LR test, and, if the null is rejected, for fixed effects using the Hausmann test. Depending on the results of these tests appropriate models have been estimated – OLS, random effects or fixed effects. These results add little to what we have seen for categories two and three (the various tests suggest that the reported models for categories four, five and six are acceptable), and the tables are not reported. Briefly: the household level random effects model is suggested for categories two and three. The only difference for group two is in the selection variables. With regard to group three, the main difference is that the farm growth variable is now significant, suggesting that on growing farms hours of off-farm agricultural labour increase (which supplements the results for this variable from Table A 2). In this case, the selection variable for group one is significant, suggesting that those in that category suited to being operators do less hours of paid farm wage labour than others.

Discussion of hours results

Reduced form hours results can be no more than suggestive. In most cases it would appear that sample selection is not a big issue, and that difficulties to explain troubling changes of sign between hours and participation estimations are uncommon. The one major change of sign that we have seen is that on education in the multinomial setting, where generally positive or zero signs for the non-agricultural categories in the individual participation equation (Table 4-7) are complemented by generally negative significant ones in the individual hours equations (Table 4-A.6). The most likely interpretation is that education encourages participation into higher earning categories, hours at work then fall because of backward bending labour supply (found in Peru by Lazlo, 2002). But the hint of a positive coefficient in Table 4-1 suggests that some of the complications dealt with by Lazlo are indeed important.

Generally, the hours estimations either complement or confirm the participation results. The variables which vary most from the participation estimations are unexpected unearned income, the availability of which clearly reduces household hours in off-farm employment (although it is insignificant in the individual estimations – which is probably due to difficulties of assignation), the number of children under seven years old, which increase individual hours in wage work (for mainly operators and adult children), and the expanding farm variable, which may be endogenous, driven by higher hours in well-paid self-employment.

Table 5-A 1: Test for linearity of self-employed returns

	Autarky		Farming + self-employment	
Dependent variable; on-farm hours	**Selection into**	**On-farm hours**	**Selection into**	**On-farm hours**
Age	-0.008	-31.124	0.003	-23.398
	[0.21]	[0.17]	[0.73]	[0.50]
Gender (1=male)	-0.050	1129.399	0.311	319.363
	[0.84]	[0.24]	[0.28]	[0.75]
Years of education	-0.077	-2.490	-0.004	44.860
	[0.00]	[0.98]	[0.83]	[0.60]
Years experience on-farm	0.015	26.415	-0.005	60.699
	[0.00]	[0.20]	[0.45]	[0.02]
Farm tenure (higher means less secure)	-0.072	-112.677	-0.024	165.657
	[0.07]	[0.57]	[0.53]	[0.19]
Farm capital	0.000	-0.198	0.000	1.260
	[0.39]	[0.32]	[0.11]	[0.00]
Farm size	-0.022	-6.385	-0.002	-11.701
	[0.29]	[0.86]	[0.92]	[0.62]
Number of children (under 15)	-0.012	12.188	0.048	-184.863
	[0.68]	[0.92]	[0.13]	[0.13]
Number of adult family members (15-65)	-0.057	888.693	-0.034	994.838
	[0.14]	[0.00]	[0.43]	[0.00]
Number of people over 65 in the house	0.117	503.819	-0.057	730.816
	[0.29]	[0.18]	[0.69]	[0.20]
Unearned income	0.000	-0.084	0.000	0.908
	[0.77]	[0.77]	[0.18]	[0.08]
Maize	-0.051	-697.826	-0.489	1324.351
	[0.79]	[0.41]	[0.03]	[0.16]
Potato	-0.088	-1168.833	0.016	-730.768
	[0.53]	[0.04]	[0.92]	[0.25]
Rice	-0.105	-912.142	0.110	347.415
	[0.70]	[0.41]	[0.64]	[0.66]
Banana	0.148	-237.697	-0.148	247.918
	[0.31]	[0.67]	[0.41]	[0.74]
Coastal region	0.259	475.271	-0.232	1513.608
	[0.18]	[0.57]	[0.33]	[0.10]
Rainforest region	0.138	-1.871	0.042	64.034
	[0.47]	[1.00]	[0.84]	[0.93]
Spanish speaking	-0.090	-481.088	0.009	-1320.292
	[0.48]	[0.37]	[0.95]	[0.01]
Inverted Simpson Index	0.075	-1.291	-0.058	160.988
	[0.13]	[1.00]	[0.29]	[0.42]
Proportion of farm output sold on the market	0.164	-361.751	-0.281	1246.506
	[0.37]	[0.62]	[0.15]	[0.12]
Unpaid labour exchange is customary in the community	0.233	-37.398	0.159	-378.779
	[0.09]	[0.95]	[0.31]	[0.53]
Average distance to amenities	-0.001		0.002	
	[0.21]		[0.03]	
Years lived in the community for newcomers	-0.001		0.003	
	[0.81]		[0.23]	
Total amount of years lived in cities by householders	0.070		-0.333	
	[0.78]		[0.25]	
Home owners	0.231		-0.065	
	[0.23]		[0.71]	
'Blow-in'	-0.269		0.073	
	[0.14]		[0.69]	
Farm size (1991)	0.016		0.007	
	[0.44]		[0.69]	
Importance of local non-agricultural income sources	-0.545		-0.028	
	[0.18]		[0.94]	
Difficulty of access (time)	0.017		-0.030	
	[0.39]		[0.19]	
Local daily rate for agricultural labour (male)		-30.578		-17.175
		[0.46]		[0.71]
Intercept	0.027	3607.363	-1.117	1906.810
	[0.97]	[0.13]	[0.08]	[0.49]
Lambda	-73.20823		-1631.778	
	[0.96]		[0.17]	
N	807	173	873	107
Wald $chi^2(43)$	142.8		136.99	

Table 5-A 2: Test of separability in off-farm work for households with just one wage earner

	On-farm hours for one-wage earner households
Age	-46.137*** [0.01]
Gender (1=male)	577.192 [0.41]
Years of education	-54.026 [0.16]
Years on-farm experience	59.659*** [0.00]
Farm tenure (higher means less secure)	-12.812 [0.86]
Farm capital	0.000 [1.00]
Farm size	39.846* [0.06]
Number of children (under 15)	-76.553 [0.24]
Number of adult family members (15-65)	521.012*** [0.00]
Number of people over 65 in the house	485.510 [0.14]
Unearned income	0.253 [0.36]
Rice	-580.456 [0.34]
Maize	-337.058 [0.25]
Banana	-399.792 [0.43]
Potato	240.621 [0.47]
Coastal region	-585.817 [0.16]
Rainforest region	-303.872 [0.44]
Spanish speaking	-633.593** [0.02]
Inverted Simpson Index	244.072** [0.03]
Proportion of farm output sold on the market	186.252 [0.64]
Importance of local non-agricultural income sources	-1,365.697** [0.04]
Difficulty of access (time)	56.580 [0.16]
Average distance to amenities	2.621 [0.24]
Average household hourly wage for off-farm wage work	-65.477 [0.53]
Intercept	3,353.055*** [0.00]
N	239
R^2	0.36

Table 5-A 3: Multinomial logit for 8 individual regimes

	On-farm work + off-farm wage work	On-farm work + off-farm wage work+self.emp.	Off-farm wage work plus off-farm self employment	Off-farm wage work only	Self-employment only	Housework only	On-farm work and self-employment
Average education of others in the house	-0.040** [0.05]	0.042 [0.51]	-0.230** [0.01]	-0.073 [0.21]	-0.066 [0.28]	-0.094* [0.09]	-0.004 [0.89]
Average gender of others in the house	-0.315 [0.15]	-1.257* [0.08]	0.593 [0.59]	-0.764 [0.27]	0.046 [0.95]	0.366 [0.57]	-1.082*** [0.00]
Average age of other adults in the house	-0.027*** [0.00]	-0.023 [0.39]	-0.062* [0.09]	0.000] [1.00]	0.001 [0.98]	-0.017 [0.46]	0.008 [0.41]
Average on-farm experience of others in the house	-0.000 [0.54]	0.029* [0.10]	0.021 [0.48]	0.009 [0.66]	0.005 [0.80]	0.017 [0.36]	-0.020** [0.02]
Age	0.122**]* [0.00]	0.189** [0.04]	0.063 [0.54]	0.193*** [0.00]	0.103 [0.12]	-0.149*** [0.01]	0.077** [0.03]
Age^2	-0.001*** [0.00]	-0.002* [0.07]	-0.001 [0.60]	-0.002*** [0.01]	-0.001 [0.22]	0.002*** [0.00]	-0.001** [0.05]
Gender (1=male)	1.182*** [0.00]	1.622*** [0.00]	-0.731 [0.23]	-0.216 [0.58]	-2.336*** [0.00]	-2.540*** [0.00]	0.519** [0.02]

	On-farm work + off-farm wage work	On-farm work + off-farm wage work+self.emp.	Off-farm wage work plus off-farm self employment	Off-farm wage work only	Self-employment only	Housework only	On-farm work and self-employment
Non-main person	-0.513**	-1.262	-0.321	1.219**	-0.553	-0.883*	-1.712***
	[0.03]	[0.13]	[0.69]	[0.03]	[0.32]	[0.08]	[0.00]
Years of education	0.041**	-0.003	0.065	0.145***	0.095**	0.059	0.047**
	[0.02]	[0.96]	[0.30]	[0.00]	[0.03]	[0.12]	[0.05]
Years on-farm experience	-0.015***	-0.036**	-37.043	-1.880***	-36.635	-3.816***	-0.015**
	[0.01]	[0.02]	[1.00]	[0.00]	[1.00]	[0.00]	[0.02]
Farm tenure (higher means less secure)	0.069*	0.112	-0.196	0.037	0.091	-0.118	-0.004
	[0.07]	[0.28]	[0.28]	[0.73]	[0.41]	[0.25]	[0.94]
Farm capital	-0.000***	-0.001**	0.000	0.000	0.000	0.000	-0.000***
	[0.00]	[0.05]	[0.47]	[0.28]	[0.22]	[0.95]	[0.00]
Farm size	-0.011	-0.02	0.05	-0.005	0.005	0.022	-0.004
	[0.25]	[0.42]	[0.16]	[0.88]	[0.87]	[0.37]	[0.71]
Number of children (under 15)	0.017	-0.157	-0.069	0.079	-0.093	0.009	0.032
	[0.61]	[0.13]	[0.66]	[0.40]	[0.37]	[0.92]	[0.48]
Number of adult family members (15-65)	-0.012]	0.273**	-0.096	0.000	0.131	0.087	0.097
	[0.79	[0.05]	[0.59]	[1.00]	[0.26]	[0.41]	[0.13]
Number of people over 65 in the house	0.148]	-0.259	-0.648	-0.746**	-0.594	-0.803**	0.049
	[0.33]	[0.64]	[0.31]	[0.05]	[0.12]	[0.02]	[0.80]
Unearned income	0.000	0.000	0.000	0.000	0.000	0.000	-0.000**
	[0.58]	[0.84]	[0.84]	[0.67]	[0.43]	[0.94]	[0.04]
Rice	-0.595***	-0.945	-43.616	-0.216	-0.124	0.054	-0.914***
	[0.01]	[0.14]	[1.00]	[0.67]	[0.82]	[0.91]	[0.00]
Maize	0.076	0.201	1.08	0.899*	1.456***	0.926*	0.294
	[0.64]	[0.71]	[0.18]	[0.07]	[0.01]	[0.05]	[0.19]
Banana	0.221	0.755	3.189***	1.661**	1.649*	1.715**	1.002***
	[0.49]	[0.27]	[0.00]	[0.04]	[0.05]	[0.03]	[0.00]
Potato	-0.366**	0.161	1.114	0.420	0.614	-0.087	0.022
	[0.03]	[0.78]	[0.17]	[0.43]	[0.28]	[0.86]	[0.93]
Coastal region	0.348*	1.174*	-0.256	-0.727	-1.256**	-0.865	-0.42
	[0.10]	[0.08]	[0.78]	[0.19]	[0.04]	[0.11]	[0.18]
Rainforest region	-0.033	0.864	-1.441	-1.134**	-0.274	-1.217**	0.401
	[0.88]	[0.19]	[0.16]	[0.05]	[0.63]	[0.02]	[0.14]
Spanish speaking	0.280*	-0.332	0.424	0.167	-0.223	0.076	0.531***
	[0.06]	[0.45]	[0.51]	[0.69]	[0.61]	[0.85]	[0.01]
Inverted Simpson Index	-0.048	-0.152	-0.262	-0.458***	-0.11	-0.176	-0.271***
	[0.40]	[0.40]	[0.34]	[0.01]	[0.53]	[0.26]	[0.00]
Proportion of farm output sold on the market	-1.315***	-1.332**	-0.138	0.264	1.277**	0.749	-1.246***
	[0.00]	[0.03]	[0.87]	[0.64]	[0.04]	[0.18]	[0.00]
Importance of local non-agricultural income sources	0.941*	1.801**	2.486*	2.301***	2.102**	1.632*	1.796***
	[0.01	[0.02]	[0.05]	[0.01]	[0.03]	[0.07]	[0.00]
Difficulty of access (time)	-0.027]	-0.06]	-0.483**	-0.058	-0.019	-0.06	-0.002
	[0.25]	[0.35]	[0.02]	[0.32]	[0.74]	[0.28]	[0.95]
Average distance to amenities	-0.002]	0.003	0.000	-0.001	0.000	0.000	0.001
	[0.12]	[0.13]	[1.00]	[0.84]	[0.98]	[0.94]	[0.17]
Intercept	-2.499***	-8.931***	1.228	-4.486**	-2.438	5.692***	-4.162***
	[0.00]	[0.00]	[0.70]	[0.03]	[0.26]	[0.00]	[0.00]
N	3002	3002	3002	3002	3002	3002	3002

Table 5-A 4: Heterogeneous households in constrained and unconstrained regimes

Dependent variable	Probability of Being Constrained	Household On-Farm Labour	Household On-Farm Labour (constrained group)	Household On-Farm Labour (unconstrained group)
Off-farm individual wage		-95.082	-378.661***	90.086
		[0.34]	[0.00]	[0.15]
Age		-37.487**	-57.318***	-8.010
		[0.02]	[0.00]	[0.39]
Gender (1=male)	0.666***	1012.469*	-1450.820**	1845.700***
	[0.00]	[0.08]	[0.03]	[0.00]
Years of education	-0.071***	-42.284	-21.578	-45.606*
	[0.00]	[0.23]	[0.55]	[0.09]
Years of on-farm experience	0.056***	59.315***	30.122**	33.447***
	[0.00]	[0.00]	[0.02]	[0.00]
Farm tenure (higher means less secure)		10.913	116.726**	-57.309
		[0.87]	[0.03]	[0.12]
Farm capital		0.319**	0.433***	0.189*
		[0.03]	[0.00]	[0.06]
Farm size	-0.039***	28.324*	148.965***	33.552***
	[0.00]	[0.10]	[0.00]	[0.00]
Number of children (under 15)		-16.319	-204.910***	
		[0.78]	[0.00]	
Number of adult family members (15-65)		566.997***	1062.060***	
		[0.00]	[0.00]	
Number of people over 65 in the house		494.158*	421.211	
		[0.07]	[0.15]	
Unearned income		0.049	0.044	-0.223
		[0.82]	[0.83]	[0.15]
Rice		-436.565	127.043	-491.281*
		[0.40]	[0.74]	[0.07]
Maize		-297.869	831.212***	-6.681
		[0.26]	[0.00]	[0.97]
Banana		-376.611	-374.734	-300.644
		[0.41]	[0.35]	[0.21]
Potato		383.931	-119.795	1209.542***
		[0.20]	[0.70]]	[0.00]
Coastal region	-0.514***	-235.860	-1257.384***	84.809
	[0.00]	[0.52]	[0.00]	[0.72]
Rainforest region	-0.349***	-20.351	-1528.753***	805.259***
	[0.00]	[0.96]	[0.00]	[0.00]
Spanish speaking		-515.329**	-1006.426***	304.850
		[0.04]	[0.00]	[0.12]
Proportion of farm output sold on the market		170.054	852.850**	-243.833
		[0.62]	[0.02]	[0.36]
Inverted Simpson Index		192.672*	-195.571**	191.294**
		[0.06]	[0.02]	[0.02]
Importance of local non-agricultural income sources		-2184.250***	-5946.626***	
		[0.00]	[0.00]	
Difficulty of access (time)		54.659	8.716	
		[0.13]	[0.80]	
Average distance to amenities		1.110	6.814***	
		[0.56]	[0.00]	
Household hours of wage work		-0.175**	-0.231***	
		[0.02]	[0.00]	
Intercept	-0.752***	3435.580***	10239.540***	481.219
	[0.00]	[0.00]	[0.00]	[0.36]
N	308	308	308	308
R^2	0.6915	0.3978	0.7996	0.3661
Adjusted R^2	0.6853	0.3444		

Table 5-A 5: Summary Statistics for constrained and unconstrained households

	Unconstrained group			Constrained group		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Household On-Farm Hours	235	2598.11	1878.37	73	3883.81	2446.51
Income Share of Off-Farm Labour	235	0.61	0.34	73	0.65	0.34
Average household hourly wage for off-farm wage work	235	1.09	1.20	73	0.62	0.39
Age	235	33.58	8.60	73	40.39	10.44
Gender (1=male)	235	0.52	0.17	73	0.59	0.21
Years of education	235	7.10	3.05	73	6.22	2.90
Years of on-farm experience	235	8.97	6.20	73	24.00	8.54
Farm tenure (higher means less secure)	235	2.08	1.72	73	1.63	1.55
Farm capital	235	518.90	778.28	73	628.75	728.52
Farm size	235	3.11	7.04	73	2.49	4.73
Off-farm work in public sector (=1)	235	0.28	0.45	73	0.34	0.48
Off farm work insured (=1)	235	0.11	0.32	73	0.05	0.23
Money spent on eating out	235	0.05	0.26	73	0.07	0.31
Number of children (under 7)	235	1.54	1.25	73	1.25	1.30
Number of children (under 15)	235	2.66	1.79	73	2.53	2.08
Number of adult family members (15-65)	235	3.34	1.60	73	3.01	1.50
Number of people over 65 in the house	235	0.16	0.44	73	0.21	0.50
Unearned income	235	178.31	470.74	73	174.55	491.68
Rent from property	235	296.43	507.57	73	225.86	391.19
Transfers	235	247.34	626.14	73	310.08	1406.72
Importance of local non-agricultural income sources	235	1.06	0.20	73	1.00	0.00
Difficulty of access (distance)	235	2.71	1.02	73	2.53	1.00
Difficulty of access (time)	235	1.71	2.99	73	1.66	2.51
Coastal region	235	0.19	0.39	73	0.07	0.25
Rainforest region	235	0.28	0.45	73	0.11	0.31
Spanish speaking	235	0.61	0.49	73	0.19	0.40
Population of survey 'segmento'	235	226.37	220.43	73	187.64	193.96
Household hours of wage work	235	1737.87	1567.90	73	1270.03	1266.61
Off-farm work in the house is seasonal (=1)	235	0.21	0.41	73	0.14	0.35
Time taken to get to off-farm work	235	9.62	18.54	73	9.76	42.78
Hours of farm labour hired in	235	0.00	0.00	73	0.00	0.00
Inverted Simpson Index	235	2.09	1.06	73	2.40	1.05
Number of household off-farm income sources	235	0.68	0.38	73	0.73	0.39
Total household income	235	5908.51	4594.69	73	3957.93	3143.53
Income per capita	235	1025.88	764.07	73	842.66	839.89

Multinomial Participation Model (for Table 5.2)

The same kinds of reasons that can constrain people's choices when they are already participating in a market (transactions costs and rationing) can also prevent people from participating in a market. The kinds of transaction costs that can have this effect are usually modelled as being either fixed or proportional (Key *et al.* 2000; de Janvry and Sadoulet, 2004), unlike the non-linear costs of the last section. A high fixed cost of entry is likely to keep some people out of the market. Only people who stand to gain enough to cover the fixed cost will participate. Proportional transaction costs raise the cost of participation by a fixed percentage for each unit (of whatever is being bought or sold – in this case, working time), and so will increase unit costs for participants. Key *et al.* (2000) consider search, bargaining and labour supervision costs as being fixed, while they consider transportation and information costs as being proportionate. Their classification is heuristic rather than rigorous, and has been disputed (Henning and Henningsen, 2005).

While we have no evidence for these costs in this study, considering the results of the multinomial regression used for the selection part of the models in the last section can be suggestive. In Table 5-A-5 below we can read off the main characteristics of farms and households in each of the four labour market regimes. Households that hire in (and do not work off-farm), for example, tend to have more females, to be older, on average, than households in autarky, to have a higher level of education and to have larger, better capitalised farms. They also tend to be in more agricultural areas, to grow rice and to be Spanish speaking households. Households that work off-farm (and do not hire in) tend to have smaller, less capitalised farms, to be nearer a range of facilities and to have a larger family size than households in autarky. They also tend to have less on-farm experience and to have weaker farm tenure. Households that both hire in and work off-farm tend to share the characteristics of households in the previous two regimes.

Table 0-A-6: Multinomial logit for 4 household work regimes (for Table 5.2)

	Hiring-In Only	Working Off-Farm Only	Both Hiring-In and Working-Off
Age	0.049***	-0.019*	0.003
	[0.00]	[0.10]	[0.78]
Gender (1=male)	-1.297***	0.847*	-0.693
	[0.01]	[0.08]	[0.13]
Years of education	0.145***	0.033	0.166***
	[0.00]	[0.37]	[0.00]
Years on-farm experience	-0.009	-0.030***	-0.017
	[0.29]	[0.00]	[0.15]
Farm tenure (higher means less secure)	0.005	0.107*	0.126*
	[0.94]	[0.09]	[0.05]
Farm capital	0.001**	-0.000**	0.00
	[0.04]	[0.02]	[0.64]
Farm size	0.069*	-0.03	0.035
	[0.10]	[0.44]	[0.32]
Number of children (under 15)	-0.02	-0.01	-0.03
	[0.81]	[0.77]	[0.58]
Number of adult family members (15-65)	-0.086	0.208***	0.141*
	[0.32]	[0.01]	[0.07]
Number of people over 65 in the house	-0.758***	-0.044	-0.352
	[0.01]	[0.84]	[0.20]
Unearned income	0.000	0.000	0.000
	[0.17]	[0.88]	[0.30]
Rice	1.272***	-0.652	0.924**
	[0.00]	[0.12]	[0.03]
Maize	0.19	0.04	0.07
	[0.56]	[0.88]	[0.85]
Banana	-0.34	0.24	-0.40
	[0.60]	[0.53]	[0.54]
Potato	0.08	-0.30	0.30
	[0.85]	[0.32]	[0.44]
Coastal region	-0.59	0.14	-0.13
	[0.19]	[0.72]	[0.78]
Rainforest region	-0.38	-0.13	-0.37
	[0.35]	[0.74]	[0.41]
Spanish speaking	0.00	0.02	0.05
	[0.99]	[0.94]	[0.87]
Inverted Simpson Index	0.01	-0.08	0.07
	[0.96]	[0.37]	[0.52]
Proportion of farm output sold on the market	1.569***	-1.053***	0.692
	[0.00]	[0.00]	[0.13]
Importance of local non-agricultural income sources	0.31	0.62	0.57
	[0.65]	[0.22]	[0.49]
Difficulty of access (time)	0.05	-0.03	-0.02
	[0.19]	[0.34]	[0.76]
Average distance to amenities	0.001	-0.004*	0.00
	[0.64]	[0.06]	[0.88]
Years in the community for newcomers	-0.012**	-0.001	-0.001
	[0.05]	[0.80]	[0.79]
Total number of years lived in cities by householders	0.279	0.213	0.325
	[0.48]	[0.63]	[0.50]
Age*farm size	0.090*	0.064**	0.109**
	[0.10]	[0.03]	[0.03]
Home owners	-0.384	-0.095	-0.271
	[0.28]	[0.75]	[0.38]
'Blow-in'	0.722*	0.003	0.211
	[0.06]	[0.99]	[0.59]
Farm size (1991)	-0.055	0.001	-0.038
	[0.24]	[0.98]	[0.28]
Unpaid labour exchange is customary in the community	-0.277	-0.465	-0.637*
	[0.35]	[0.10]	[0.10]
Intercept	-3.418***	0.851	-2.321*
	[0.01]	[0.41]	[0.07]
N	980	980	980

Typically, the multinomial logit model is justified by a random utility model, and column choice is assumed to be based on agents' choice of their highest utility option. In this case, however, the availability of the highest utility option may be conditional on the existing constraints, and the actual option chosen is therefore the highest *available* utility option. The variables in the multinomial logits may therefore be divided into (non-mutually exclusive) variables that affect regime-specific utility (variables, that is, that affect farm productivity, the shadow value of home time and real off-farm wage) each of which we can define as being elements of a vector – v – and variables which affect constraints on achieving a particular level of utility – the variables in the vector z, defined in equation (5).

Assigning variables to z is difficult to do in advance. From some of the results of Table 5-4 it is possible to speculate.

Higher education levels, for instance, significantly increase the chance of entry to regimes 2 and 4, and are completely insignificant for 3. This adds significantly to the explanation of the education effect seen in reduced form in Table 4-1. Having more education has a double effect – more hiring in (probably because the shadow value of home time rises and/or the intensity of farm production rises) and more off-farm work (as seen in Table 4-1), but only if hiring in also takes place, suggesting that the off-farm work for the educated is paid better than the available farm wage. Education, thus, is probably a 'v' variable; but it could affect constraints on entry also (and so be in 'z') if more education lowers search costs (both for labour to hire in and jobs to work outside in) and bargaining costs generally.

The variables associated with the farm and the human capital associated with it (farm experience, farm capital, farm tenure, the diversification coefficient, the market embeddedness coefficient, farm size and main crop outputs) are almost certainly to be exclusively in v. Gender and remoteness, on the other hand, are probably in 'z' as well as in 'v'. Remoteness appears not to affect the hiring in decision, only – as also seen occasionally in Chapter 4 – the off-farm decision. This is also affected by the minka variable. As in Table 4-1, the language variable is nowhere significant. Being non-Spanish speaking does not appear to affect either the hiring in or working off-farm decision.

As determinants of transaction costs that might hinder entry into categories 2, 3 or 4, then, it seems likely that gender, education and location are plausible candidates, if such costs exist. Language spoken appears not to be but as seen in Chapter 4 the language dummy effect does not tend to come through at household level, only at individual level.

Table 6-A 1: CLAD model using off-farm earnings, not hours

		Technical efficiency					Allocative efficiency					Scale efficiency					
	Reps	Observed	Bias	Std.Err.	[90% Conf. Interval]		Observed	Bias	Std.Err.	[90% Conf. Interval]		Observed	Bias	Std.Err.	[90% Conf. Interval]		
Spanish speaking	1000	-0.057	0.039	0.118	-0.250	0.137	0.041	-0.016	0.081	-0.093	0.175	0.083	0.057	0.124	-0.122	0.287	(N)
					-0.184	0.218				-0.113	0.154				0.020	0.350	(P)
					-0.244	0.111				-0.081	0.180				-0.057	0.191	(BC)
Inverted Simpson Index	1000	0.091	-0.009	0.047	0.013	0.169	-0.061	0.006	0.026	-0.103	-0.019	-0.032	-0.017	0.048	-0.110	0.047	(N)
					0.000	0.160				-0.093	-0.009				-0.141	0.009	(P)
					0.019	0.175				-0.101	-0.018				-0.111	0.017	(BC)
Proportion of farm output sold on the market	1000	0.844	-0.089	0.165	0.573	1.115	0.137	0.045	0.110	-0.044	0.318	-0.060	-0.034	0.137	-0.286	0.167	(N)
					0.468	0.997				0.012	0.375				-0.337	0.075	(P)
					0.684	1.308				-0.024	0.290				-0.296	0.090	(BC)
Years of education	1000	0.008	-0.007	0.012	-0.011	0.028	-0.002	0.002	0.008	-0.015	0.012	-0.004	-0.003	0.013	-0.025	0.016	(N)
					-0.021	0.018				-0.012	0.016				-0.034	0.007	(P)
					-0.005	0.030				-0.015	0.012				-0.033	0.008	(BC)
Gender (1=male)	1000	-0.218	-0.032	0.152	-0.468	0.032	0.043	0.010	0.119	-0.153	0.239	0.047	0.075	0.168	-0.230	0.324	(N)
					-0.498	-0.002				-0.139	0.253				-0.091	0.427	(P)
					-0.426	0.044				-0.161	0.229				-0.241	0.221	(BC)
Farm size	1000	-0.005	0.010	0.023	-0.043	0.033	-0.034	0.000	0.011	-0.052	-0.015	-0.037	-0.011	0.023	-0.076	0.001	(N)
					-0.027	0.047				-0.052	-0.014				-0.092	-0.016	(P)
					-0.032	0.033				-0.052	-0.014				-0.084	-0.006	(BC)
Age	1000	-0.001	0.000	0.003	-0.006	0.003	-0.002	0.001	0.002	-0.005	0.002	0.001	-0.001	0.002	-0.003	0.004	(N)
					-0.006	0.003				-0.004	0.003				-0.004	0.003	(P)
					-0.006	0.003				-0.005	0.000				-0.003	0.004	(BC)
Off-farm earnings	1000	-0.0029	-0.0016	0.0047	-0.0107	0.0049	0.0031	0.0002	0.0032	-0.0022	0.0083	-0.0008	0.0013	0.0047	-0.0085	0.0070	(N)
					-0.0132	0.0023				-0.0021	0.0083				-0.0055	0.0098	(P)
					-0.0088	0.0052				-0.0024	0.0081				-0.0091	0.0052	(BC)
Intercept	1000	0.214	0.102	0.234	-0.171	0.598	0.559	-0.102	0.174	0.272	0.846	1.015	0.051	0.237	0.625	1.404	(N)
					-0.030	0.729				0.148	0.722				0.818	1.497	(P)
					-0.188	0.536				0.340	0.852				0.831	1.555	(BC)
Initial sample size		368					368					368					
Final sample size		345					368					319					
Pseudo R^2		0.235					0.117					0.080					
N = normal, P = percentile, BC = bias-corrected																	