AIX-MARSEILLE UNIVERSITÉ
FACULTÉ D'ÉCONOMIE ET DE GESTION

École Doctorale de Sciences Economiques et de Gestion d'Aix-Marseille ED 372

Année 2014 Numéro attribué par la bibliotèque
| | | | | | | | | | | | |

Thèse pour le Doctorat en Sciences Economiques

Présentée et soutenue publiquement par

XUN ZHOU

le 16 Décembre 2014

Econometric analysis of subjective well-being, preference, perception and dynamics

---

DIRECTEUR DE THÈSE

M. Michel Lubrano
Directeur de Recherche CNRS, GREQAM
M. Emmanuel Flachaire
Professeur à Aix-Marseille Université GREQAM

JURY

RAPPORTEURS

M. Frank Cowell Professeur à la London School of Economics
M. Andrew Clark Professeur à l'Ecole d'Economie de Paris

EXAMINATEURS

M. Stephen Bazen Professeur à Aix-Marseille Universitè, GREQAM

This thesis was prepared at Aix-Marseille School of Economics, Greqam

À mes grands-parents,
À mes parents,
À ma femme SONG Yu,
À mes amis.

Acknowledgment

First and foremost, I have to thank my research supervisors, Michel Lubrano and Emmanuel Flachaire. Without their assistance and dedicated involvement in every step throughout the writing process, this thesis would have never been accomplished. I would like to thank them very much for their support and understanding. And I'm particularly grateful to Michel Lubrano who played not only the role of a tutor but also the role of a friend and a kind elder over these past four years. His immense knowledge and precious advices enlightened me in every discussion we had.

I would like to thank my examiners Professor Frank Cowell and Professor Andrew Clark for their helpful comments and corrections. Their suggestions were very important in improving the quality of this dissertation. I would like also thank particularly Professor Stephen Bazen with whom I had many usefull discussions. His help and comments were particularly welcome.

I am also very grateful that I could pass my Ph.D at GREQAM, one of the most important quantitative economic institutes. I want to thank the administrative secretaries, Mme. Bernadette Vouriot, Mme. Isabelle Mauduech and Mme. Aziza Sikar for their patient helps. I would like to thank Professor Velayoudom Marimoutou for providing me an Office in Luminy that greatly improved my efficiency. I would also like to thank Professor Christelle Lecourt who gave me the idea to write Chapter 6 of this dissertation, a paper which widely opened my views on simulation techniques in econometrics and gave rise to a probable good publication.

Last but not least, I would like to thank my family and my precious friends. I could not be any luckier having my grand-parents and parents, who have always been supportive of my projects during all these years studying in France. Their love and care for me became my rock in spirit. Sometimes I feel guilty for my absence while I believe absence makes the heart grow fonder. I wish this work could express my respects. I am grateful to have met my lovely wife Song Yu during those years, my better half, who brings happiness and provides endless understanding and love in my life. Finally, I would like to express my deep gratitude to all the people who provided help to me, all the persons that I cannot mention specifically by their names here.

Avertissement

Mise à part l'introduction et la conclusion générale qui constituent les chapitres 1 et 7, les différents chapitres de cette thèse sont issus d'articles de recherche rédigé en anglais et dont la structure est autonome. Ceci y explique la présence des termes "paper" ou "article" ainsi que l'éventuelle répétition de certaines informations.

Notice

Except for the general introduction and the general conclusion which form chapter 1 and chapter 7 of this dissertation, the remaining chapters (2 to 5) of this dissertation are self-contained research articles. Consequently, the terms "paper" and "article" are frequently used. This also explains that some pieces of information are given in multiple places.

Résumé

Cette thèse s'intéresse à l'analyse des variables d'opinion. Les opinions couvertes concernent spécifiquement des questions économiques comme le niveau de bien-être, la situation financière, le niveau minimum de revenu nécessaire pour vivre dignement, la préférence pour la redistribution. Le traitement de ces variables d'opinion et leur mise en relation avec les grandeurs Ãľ-conomiques traditionnelles comme le niveau de revenu ou sa dynamique nécessitent des techniques micro-économétriques spécifiques. Dans cette thèse, les modèles dynamiques de panel sont utilisés pour étudier la mobilité des revenu et la mobilité des statuts d'emploi. Dans le dernier chapitre, nous discutons également le modèle Tobit dynamique et l'importance des conditions initiales. Dans les modèles de panel, se pose la question du choix entre effet fixe et effet aléatoire. Parce que les variables subjectives sont ordinales et discrètes, les effets fixes sont difficilement identifiables. L'identification des effets alèatoires est moins problématique, mais l'estimation devient pourtant difficile quand la dimension d'intégration augmente. Pour résoudre cela, je l'utilise intensivement les techniques de simulation, soit pour le modèle dynamique multinomial logit, soit pour le modèle dynamique Tobit. La technique de simulation est également appliqué au modèle trivarié probit pour mesurer les corrélations conditionnelles entre trois (ou même plus) variables ordinales.

Mots clé: *bien-être subjective, préférence pour la redistribution, panel, Method simulation, perception, dynamics.*

Abstract

This dissertation studies option variables. These opinions cover specially economic questions such as the level of wellbeing, financial situations, the minimum income question and the preference for redistribution. The treatment of these opinion variables and their relation to conventional economic questions such as income level or dynamics requires the use of special micro-econometric models. In this dissertation, the dynamics panel models are used to study the job status and income mobility. In the 6 chapter, we discuss the dynamic Tobit model with an emphasis on initial conditions. The choice between fixed or random effect is another question. Because subjective variables are ordinal and discrete, the identification of fixed effects is problematic. Random effects are better identified while the estimation difficulty increases with the integration dimension. To solve that, I use intensively simulation method in the study of dynamic multinomial logit model or dynamic tobit model. It is also been applied in the trivariate probit model to measure the conditional correlations among more than 2 ordinal variables.

Key words*: Subjective wellbeing, preference for redistribution, panel data, Simulation method, perception, dynamics.*

Contents

1 General Introduction **1**

1.1 The subjective approach 2

1.1.1 Relative income and SWB 4

1.1.2 Relative Income and international poverty 5

1.1.3 Redistribution preference and POUM 7

1.1.4 Redistribution preference and perception of the origins of poverty 8

1.2 Econometric tools 9

1.2.1 Property of subjective variables and the use of the panel data 9

1.2.2 The evaluation of dynamics 11

1.2.3 Endogeneity 13

1.3 The structure of this dissertation 14

2 Reference Groups **17**

2.1 Introduction: reference groups 17

2.2 A survey of comparison income definitions 18

2.2.1 Subjective reference groups 19

2.2.2 Researcher defined reference groups 20

2.2.3 Characterising the reference group 21

2.3 Economic and econometric assumptions 21

2.3.1 Basic model 21

2.3.2 Panel data models 23

2.3.3 Panel data and income dynamics 23

2.3.4 Reference income and income inequality 25

2.3.5 Identification and likelihood function 26

2.4 An investigation using the BHPS 27

2.4.1 Income dynamics 28

2.4.2 The choice of a reference group definition 28

2.4.3 The empirical content of reference groups 31

2.4.4 The puzzle of personal versus reference income 32

2.4.5 Asymmetric effects . . . 33
2.5 The impact of inequality . . . 35
2.5.1 Inequality and reference groups . . . 35
2.5.2 Identifying risk versus opportunity . . . 36
2.6 Conclusion . . . 39

3 Bayesian subjective poverty line 41
3.1 Introduction . . . 41
3.2 Poverty lines and preference drift . . . 43
3.2.1 Subjective poverty lines . . . 43
3.2.2 Absolute poverty lines for developing countries . . . 44
3.2.3 Evidence of preference drift among developing countries 45
3.2.4 An international subjective poverty line . . . 47
3.3 Bayesian regression with a break . . . 48
3.3.1 Likelihood and posteriors . . . 48
3.3.2 The two variance case . . . 50
3.4 Data and estimation . . . 51
3.4.1 Revisiting the initial model . . . 51
3.4.2 Preference drift . . . 55
3.4.3 How to simulate the posterior density of the IPL . . . 57
3.5 Conclusion and comments . . . 59

4 Preference for redistribution and POUM 61
4.1 Introduction . . . 61
4.2 A simple model for redistribution preferences and income mobility . . . 63
4.2.1 A lognormal example . . . 65
4.2.2 The POUM effect and risk aversion . . . 65
4.2.3 Stochastic dominance and regular income dynamics . . 68
4.2.4 Heterogeneity . . . 69
4.2.5 Related empirical work . . . 69
4.3 Stylized facts from the BHPS . . . 70
4.3.1 Job status . . . 70
4.3.2 Wage mobility . . . 71
4.3.3 Attitude to redistribution . . . 73
4.3.4 Our empirical strategy . . . 76
4.4 Modelling job status mobility . . . 78
4.4.1 A dynamic multinomial logit model . . . 78
4.4.2 Initial conditions and individual effects . . . 79
4.4.3 Marginal effects . . . 80
4.4.4 Empirical job status mobility . . . 81

4.4.5 Unemployment risk, chances to go back to work 82
4.4.6 Implicit transition matrices for job status 84
4.4.7 Testing for the validity of the model 84
4.5 Wage mobility in an heterogeneous world 85
4.5.1 Wage dynamics and heterogenous Markov processes . . 86
4.5.2 Implied transition matrices and individual heterogeneity 87
4.5.3 Probability to go to the the upper quantiles 88
4.6 Preference for redistribution and income mobility dynamics . . 89
4.6.1 The working group 90
4.6.2 The unemployed group 93
4.6.3 The non-participating group 94
4.7 Conclusion 95

5 Redistribution and inequality in China 97
5.1 Introduction: Inequality in China 97
5.2 Literature review 100
5.3 The Chinese General Social Survey 102
5.3.1 Exogenous variables 102
5.3.2 Dependent variables: Social values and opinions 103
5.4 Occupation and Social mobility 105
5.4.1 Inter-generational mobility via Markov chain 106
5.4.2 Occupation mobility 107
5.5 Econometric modeling of preference for redistribution 109
5.5.1 A trivariate ordered probit model 111
5.5.2 Monte Carlo simulation 114
5.5.3 Evaluation strategy 114
5.6 Preference for redistribution in China 116
5.6.1 Structural correlation 116
5.6.2 Poverty perceptions 118
5.6.3 Preference for redistribution 119
5.6.4 Rural-urban segmentation 120
5.7 Conclusion and discussion 121

6 A two-tiered dynamic panel Tobit model 125
6.1 Summary 125
6.2 Introduction 126
6.3 Likelihood function and initial conditions 127
6.4 Checking the empirical results 128
6.4.1 Optimization strategy 129
6.4.2 Empirical results 130
6.5 Conclusion 133

7 General conclusion **137**

A Appendices **139**
A.1 Chapter 2 139
A.1.1 CASMIN levels 139
A.1.2 Metropolitan areas 139
A.1.3 Comparing two independent regressions 140
A.2 Chapter 3 141
A.2.1 Simulation of a bivariate density using a grid 141
A.3 Chapter 4 142
A.4 Markov processes 142
A.4.1 Mobility indices 142
A.4.2 Monotone transition matrices 143
A.4.3 Equality of opportunity 144
A.4.4 Testing for homogeneity 145
A.4.5 Testing for regularity or monotonicity 146
A.4.6 Testing for progressivity 147
A.5 Chapter 5 148
A.6 Properties of the Transition Matrix 148
A.6.1 Mobility indices 148
A.6.2 Proof 150
A.6.3 RNG 151

List of Tables

2.1 Estimation of a first life satisfaction equation 29
2.2 Four models of life-satisfaction using median income of different reference groups 30
2.3 Gini for educational categories 2002-2008 31
2.4 The full puzzling model 32
2.5 Estimation of an asymmetric satisfaction equation 34
2.6 Estimation of a life satisfaction equation 37
2.7 Estimation of a life satisfaction equation 38
2.8 Estimation of a life satisfaction equation 39

3.1 Bayesian inference for initial model 53
3.2 Bayesian inference the two variance case 55
3.3 Model with preference drift and two variances 56

4.1 Percentage of a POUM effect 67
4.2 Mean number of observations per year for each job status . . . 71
4.3 Average yearly transition matrix between job statuses 71
4.4 Yearly income transition matrices 72
4.5 Probabilities to move to (or stay in) 73
4.6 preference for redistribution in the UK 74
4.7 Mobility among preference for redistribution 75
4.8 Preference for redistribution as a function of wage quantiles . 75
4.9 Estimation of a dynamic Multinomial Logit 82
4.10 Implicit conditional transition matrices 84
4.11 Wage quantile dynamics with observed and unobserved heterogeneity 87
4.12 Marginal effects for wage quantile dynamics 88
4.13 Redistribution preference when in a working spell 91
4.14 Redistribution preference for unemployment spells 93
4.15 Redistribution preference for out of the labour market spells . 94

5.1 Socio-economic descriptive statistics 104
5.2 Income distribution 105
5.3 Attitude and perceptions 105
5.4 Inter-generational mobility, Prais Index = 0.774 107
5.5 A monotonicity test 107
5.6 Inter-generational occupation mobility 110
5.7 Mean bias for comparing three independent ordered probit . . 115
5.8 MSE for comparing three independent ordered probit 115
5.9 Preference for redistribution 117
5.10 attitudes of migrant workers and new urban residents 122

6.1 Correlated RE+AR(1) one-tiered dynamic panel Tobit model . 134
6.2 Correlated RE+AR(1) two-tiered dynamic panel Tobit model 135

A.1 CASMIN lelvels, last wave 139
A.2 Metropolitan areas, last wave 140

List of Figures

3.1 Why a model in logs is better 52
3.2 Rescaled posterior density of α_1 and posterior density of θ . . 53
3.3 Posterior fit of the initial model in logs 54
3.4 Posterior density of θ and posterior fit of the last model 56
3.5 Posterior density of four poverty lines 58

4.1 A recursive empirical model to explain 76
4.2 Predicted job status probabilities 83
4.3 Probability to move to or stay in the upper categories (Q4-Q5) 89

Chapter 1

General Introduction

Social welfare is one branch of economics that evaluates collective well-being. The welfarist approach concerns mainly the distribution of observed welfare and its implications while well-being is by nature multi-dimensional. The concerns of individuals may not be simply what they have but where they are on the social ladder. Moreover, future prospect, perceptions and psychology traits may have important impacts. As underlined by Sen, the perceived well-being is highly heterogeneous among individuals, and even if resources are equally distributed, the level of well-being as perceived by the different individuals would not be the same. The subjective approach comes after the welfarist approach and became a very important methodology in the last decades.

This branch of study investigates directly how people perceive their current or future status and what are their attitudes or values towards these different questions. Departing from the traditional view of revealed preferences, the subjective approach assumes that individuals are experts of themselves and they report what they feel in a consistent way. In survey data, one typical "subjective well-being" (SWB) question is the following: *How dissatisfied or satisfied are you with your life overall?*. Respondents answer to this kind of question by choosing a number from a discrete monotone ordinal scale, e.g. 1 for totally dissatisfied till 5 for totally satisfied. This is a Likert scale based on 5 items. This question, which is also called the "Happiness" or "Satisfaction" question, is the most used one in the subjective approach. Since the aim of economic growth is to improve the living standard of human-beings, making people happier seems to be one of the ultimate goals. Some other domain satisfaction questions could also be used to access the quality of life in different aspects, such as job satisfaction, family satisfaction, etc. The subjective approach is a multidisciplinary study field which is not only being used by economists, but also by psychologists, sociologists and politicians. The subjective approach does not limit the lens to the satisfaction questions, but also to the subjective attitudes, preferences, perceptions as reported by people.

1.1 What could we learn from the study of subjective approach?

Although some researchers have doubt on the reliability of the answers of these kind of questions (even the reliability of the approach), empirical studies covering many countries over many years provide clear evidences that people behave in a rational way. For a review of the use of subjective data sets see for instance Frey and Stutzer (2002). Since then, researchers have begun to use these data to answer a variety of questions. For instance, what is the determinant of SWB, which is the most discussed issue in Happiness economics. Income, of course, as an indicator of the access to social resources in a given society, should play an important role in SWB. However, SWB is not a simple function of income. As the wealth of the society cumulates, the perception of income should also vary. Other factors also have important effects upon SWB, such as education, type of job, health status, family life, etc. However, "happiness" or "life-satisfaction" are very vague concepts, many important factors are unobservable, such as psychological traits (optimism or pessimism), anchoring scales (how people understand the word "happiness"), culture, beliefs and life-cycle experiences, etc. How to describe the SWB then becomes a challenge.

Knowing the determinant of SWB is not the ultimate goal. It was found that in firms, a higher level of SWB is associated with a higher level of human performance, see for instance Oswald et al. (2009). But the causal link is not clear enough because the premium of higher human performance could also lead to higher well-being. Obviously, almost all subjective factors are mutually interrelated, and the causality is far too complex to be understood. For example, the endogeneity relation between over all satisfaction and domain satisfaction deserves more attention for researchers. van Praag and Ferrer-i-Carbonell (2004) provide one possible way to solve the endogeny problem between over-all and sub-domain satisfactions. Another important issue of interest for the economists is the interpersonal or international comparability of satisfaction levels.

The subjective approach also opens another possibility of assessing the effect of policies. For example, in 1972, American senator George McGovern's proposals for a tax reform and an income redistribution scheme was rejected by the majority of Americans. Knowing the stylized fact that in most (*or* all) developed countries, the mean level of income is higher than the median income, the silence of the majority and its opposition to redistribution is then puzzling. The explanation of this issue might need to go back to long term history (see for instance Alesina and Glaeser 2004). But it is also possible to find that the majority of the American population actually reacts according to a particular rationality. In order to shed light on this hidden rationality, Benabou and Ok formalised the hypothesis called "prospect of upward mobility" (POUM) in their famous paper Benabou and Ok (2001c). In this paper, the focus is on the income dynamics recorded in the past and the revealed voting preferences. Following them, by using subjective information collected from different data sets in different countries, researchers found empirical evidence of the "POUM" hypothesis. See for example the studies of Clark and d'Angelo (2008a) and Idema and Rueda (2011) with the British Household Panel Survey (BHPS) data set; Alesina and La Ferrara (2005) with the International Social Survey Program (ISSP) data set, Schwarze and Harpfer (2007) with the German Social Economic Panel (GSOEP) data set and Xu and Liu (2013) with the Chinese General Social Survey (CGSS) data set, Ravallion and Lokshin (2000) with the Russian Longitudinal Monitoring Survey (RLMS). In all these data sets, respondents anchor their attitudes on a Likert scale. Their answers are directly or indirectly related to their preference for redistribution. Empirically, the "POUM" hypothesis requires the measurement of social mobility. The mobility measurement has also been used in SWB studies, but mostly for backward mobility (hedonic adaptation, see for instance Lyubomirsky 2011, Shane and Loewenstein 1999 and Kahneman et al. 1999). Instead, the "POUM" hypothesis focuses on forward income mobility rather than on social status at given point of time. Then the prospect of mobility enters into the utility function used for reporting people preferences. The prospect of upward mobility could be measured by income mobility as in Alesina and La Ferrara (2005) or occupation mobility as in Clark and d'Angelo (2008a). All evidence converge to one fact that people who experienced or anticipated an upward mobility are less in favour of redistribution (and also happier). However, the mobility perceived by individuals whose current status is lower or greater than the mean level should not be the same, just as suggested by the "POUM" hypothesis. This question is less discussed in the empirical literature. Meanwhile, risk aversion should also be an important control variable as mentioned in Benabou and Ok (2001c). People have different attitudes towards redistribution not only because they held prospects of future gains, but also because they might appraise differently the security of what they already have. Another specific approach is called the "'tunnel effect" (Hirschman and Rothschild 1973) which focuses on the effect of the dynamic experience of others around, as perceived by the respondents. Similarly to the relative income hypothesis, a comparison effect plays a very important role, see for instance the evidence found in Ravallion and Lokshin (2000) using a Russian data set.

Other than the self-interest point of view, there are still many unobserved factors that enter the preference utility function. This leads to another strand of preference studies that relates individual preferences to psychological traits, values, beliefs and so on. One of the interesting question is the role of perception of origins of poverty played in the forming of the preference for redistribution. As found in Alesina and Glaeser (2004), US people tend to be less in favour of redistribution while EU people are more in favour of it because people who come from the two sides of the Atlantic held different perceptions of what is the origin of poverty (also called justice recognition, see e.g. Xu and Liu 2013). These authors call this relationship the "sense of justice". This is saying that people have the common sense that one should hold responsibility of his own choices while he's innocent for what comes to him out of his own control, just as argued by Rawls (1971). Especially, "efforts" and "circumstances" are the two factors mostly considered under the "sense of justice".

In the next few subsections we briefly introduce the topics that we have surveyed in this dissertation.

1.1.1 Relative income and SWB

The relationship between SWB and the level of income is one of the most discussed issue in this field. As was first found by Easterlin (1974), one empirical puzzle is that at a given point of time and for a given country, richer people are happier than poorer ones, but when time passes an increase in GNP does not correspond to an increase in average happiness. Several explanations were given to this paradox (see for instance the survey of Clark et al. 2008). Obviously, satisfaction includes basic human needs, but once these basic needs are met, individuals start to be concerned by what others have. Researchers then started to assume that happiness or utility is not a direct function of income, but a function of the relative income (or the distance between individual income and the mean income of the reference group, the latter being defined as "people like me"). For empirical evidence see for instance Ferrer-i-Carbonell (2005), Blanchflower and Oswald (2004), Clark and Oswald (1996), Easterlin (1974) and Ferrer-i-Carbonell (2005). A higher status brings in positive effects for subjective well-being while a relative low status brings in negative effects. On the other hand, how people perceive their reference group is not simply how they compare themselves to others, but also how they anticipate their future position from examples of "people like me". Obviously, it can be a sign of opportunity or of risk. How to identify the effect of perceived opportunities and the effect of perceived risks is a question that remains unsolved. Using the BHPS data set, we found that the usual theory of adaptation (individuals get used to their income level and react only to variations of it, see Clark et al. 2008) is associated to the comparison effect and that the choice of the reference group is not very sensitive. We have also discussed some specific puzzles found in the BHPS. The first puzzle is that the mean or median level of reference income has much higher effect than the permanent income. This means that the relation between SWB and reference income is non-linear. If the position does not change, well-being decreases with long term income. We then have to dig more information provided by the reference group, i.e. the dispersion. The dispersion of the reference income is measured by a Gini index. It has been found that the a higher value of the Gini index in the reference group is associated with higher SWB which comes as another puzzle. We finally manage to solve the two puzzles by introducing the Gini index of each region where respondents live in which has a significant negative effect. This means that British people perceive a variety of information from the reference group and the reference group is not unique. The gap between individual income and the reference income has a negative effect while individuals also perceive opportunity or risk from the dispersion of a reference income distribution, depending on their social status.

1.1.2 Relative Income and international poverty

The relative income theory can also be applied to determine a poverty line. The basic assumption is that, between different countries, the minimum basket of goods that ensures physical and mental well-being is not the same, just because living standards, traditions, habits and other social characteristics are different. With the development of human society, the critical level under which people are recognized as being poor also varies. Thus the definition of poverty is no longer the state of being starving. Assuming that basic needs are met, except in the least developed areas, poverty means that people are unable to be like the others and do not manage to take a decent part in social life. By definition, the necessary level that ensures basic needs is called "absolute poverty line" such as enough food, clean water, sanitation, clothing, shelter, health care and basic education. Beyond that, "poverty" becomes more complex and is influenced largely by the perception of "economic inequality". Under this context, we turn to the notion of "subjective poverty". A "subjective poverty line" is the minimum level of income required to lead a decent life as subjectively measured by individuals. For example, the approach based on the "minimum income question" (MIQ, see e.g. Kapteyn et al. 1988). Or, the approach based on the "Income Evaluation Question" (see e.g. van Praag and Ferrer-i-Carbonell 2004). But the perceived levels of income for different levels are heterogeneous not because people are heterogeneous, but also because of the different levels of actual income they have. At a country level, an official poverty line denotes a level of income that is commonly being consider as the necessary minimum by the welfare state. Ravallion et al. (1991) showed that official national poverty lines varied little in comparison with mean consumption per capita for less developed countries, while above a critical level of mean consumption per capita, national official poverty lines had a much stronger elasticity with respect to consumption. Clearly, preference drift (van Praag and Ferrer-i-Carbonell 2004) plays an important role in determining "relative poverty".

Based on that previous finding, Ravallion and Chen (2001) and Ravallion and Chen (2004) proposed an international poverty line (a worldwide absolute poverty line) as being "$1 per day" ($1.08 at 1993 PPP). In a more recent paper, Ravallion et al. (2009) clearly identify two groups of countries in a new data set covering 74 developing countries with data collected within the period 1988-2005. They estimate a non-linear regression relating national official poverty lines to national mean consumption level per capita, imposing a zero consumption coefficient for the group of less developed countries. With this model, the revised International Poverty Line (IPL) has risen to $1.25 per day at 2005 PPP. Greb et al. (2011) revisited this study, using different econometric techniques and a different specification and found a somewhat higher international poverty line at $1.45 per day. The huge difference between $1.25 and $1.45 means that 317.6 million supplementary people would fall in poverty in 2005. The poverty lines reported in both Ravallion et al. (2009) and Greb et al. (2011) lack a good precision. Moreover, the watershed between "absolute poverty" and "relative poverty" is defined *a priori* by a breakpoint between zero consumption elasticity and a positive consumption elasticity. The zero consumption elasticity in the least developed countries is an assumption that should be checked. Without doubt, the study of the IPL would be very useful to find the location and the amount of people who are in poverty, which deserves further examination. We found a new IPL equals to 1.48 dollar a day which is higher than the one suggested in Ravallion et al. (2009). The definition of the IPL has also changed. Even within the least developed countries, we still found a small but significant consumption elasticity. Although, the consumption elasticity for the lowest part is much lower than countries located on the right hand side of the breakpoint. Because the data set has only 74 observations, our Bayesian breakpoint regression performs better than the one used by Ravallion et al. (2009) and Greb et al.

(2011). Additionally, we also found that the estimation errors for the left group (below the breakpoint) and the right group (above the breakpoint) are heterogeneous. Our final results fit better the data while use much narrower standard intervals.

1.1.3 Redistribution preference and POUM

The study of preferences for redistribution is mostly conducted over two strands. The first strand of studies is based on the self-interest point of view. Benabou and Ok (2001c) gives the theoretical linkage between individual preferences for redistribution and the prospection of mobility. They also argue that the effect of prospect of mobility is associated with risk aversion. Since mobility prospect is based on the observations or experiences of occurred mobilities, the measurement of mobility is then one core question in the study of preferences for redistribution. Conditional (heterogenous) mobility is more interesting than unconditional (homogeneous) mobility because it provides the specified mobility pattern of individuals for different groups. The measurement of risk aversion is also important while it is not an easy task. Under globalisation, high skilled individuals face greater opportunities while the low skilled individuals are risking losses due to the competition with low skilled workers from less developed countries. People who have different characteristics should have heterogeneous prospects of their own mobility, either an opportunity or a risk, or both of them. The first possible way to measure the opportunity/risk level is to investigate the dynamics of job status. Another way is to illustrate the income dynamics. Both aspects are important in that the income dynamics is always associated with changes in job status. On the other hand, we could extract different pieces of information from the two dynamics. For example, having a job implies more risk of loss compared to unemployed people simply because the latter have nothing more to lose. However, individuals who do not have a job or do have a job which is not highly paid anticipate greater improvements if the overall dynamics is progressive and concave. How do different mobilities affect the preference for redistribution is then a question. Using the BHPS data, we estimate the relationship between preferences for redistribution and job/income dynamics. The dynamic characteristics are extracted from the full length of the panel (18 waves). We estimate the dynamic process by dynamic multinomial logit models with observed variables and random effects and then use the dynamic parameters to predict the future expected status of each individual according to their current social status. We found that depending on the job status, the determinants of the preference for redistribution are heterogenous. Within the working group, we found a significant positive "POUM" effect.

The "POUM" is also found to have an asymmetric effect in that individuals having current income lower than the median level respond in a stronger way to the "POUM" effect. Individuals in the working group are also very sensitive to the risk of unemployment. The unemployed group behaves in a different way. They are more sensitive to the probability of getting a job and the effect of expecting better financial situation (subjectively) is also higher. Individuals in the non-participating group (marginal group) are mostly women. The most important factor found for this group is the household income mobility.

1.1.4 Redistribution preference and perception of the origins of poverty

The second strand of studies on preferences for redistribution is based on the idea that preferences not only reveal the preferences for social status, but also attitudes and values depending on complicated mechanisms. It is hard to illustrate the causality between preferences and values/attitudes, but the correlations can possibly be evaluated. We might find thousands of subjective factors that are correlated with preferences, among them, researchers are mostly interested by the perception of the "sense of justice". In the literature (see e.g. Alesina and Glaeser 2004 and the references quoted there), we have clear evidence of the linkage between preferences and "sense of justice". They are based on surveys while the magnitude of the correlations are less surveyed. Thus the preference for redistribution mechanism deserves further investigations. It is then natural to suppose that these subjective variables interplay altogether in a simultaneous system. By exploring the CGSS, we found two questions concerning the "sense of justice":

1. Do you agree that individuals are poor because society is not well functioning, especially because of misgovernment?
2. Do you agree that individuals are poor because they are lazy?

Answers to these questions (on an ordinal scale) reflect the respondents' perceptions about the origins of poverty, which are potentially correlated to the preferences for redistribution. Then we model these three ordinal variables (including the preference for redistribution variable) using a tri-variate ordered probit model which we estimated via a Geweke-Hajivassiliou-Keane (GHK) simulator. According to a Monte Carlo experiment (sample size of 1 000 with 1 000 replications), the proposed model converges quickly and stably. The estimation with the real data tells us several important stories. First, the three subjective ordinal variables are correlated after controlling for the same set of explanatory variables. People who agree that individuals are poor because of misgovernment are more in favor of redistribution, and the contrary for laziness. However, we found that misgovernment and laziness are not mutually exclusive. This is evidence that at least for ordinary Chinese people, "effort" and "circumstances" are both important factors in term of success and poverty. Note that "misgovernment" plays a much stronger role in people's perception. The model proposed in this study also provides the possibility to survey the direct and indirect effect of the explanatory variables. Evidence of the "POUM" effect has also been found in the CGSS data.

1.2 Methodologies and econometric tools

In this section, we discuss briefly the use of subjective data and some basic econometric methodologies applied in this field. The subjective approach, especially Happiness economics, gives the impression that econometrics techniques rely on some simple linear regressions and straightforward interpretations. However, the human thought is complex, the subjective variables usually provide very brief or vague information. Several assumptions and adjustments have to be made in order to perform the analysis, which leads to many improvements in both economics and econometrics. In particular the panel dimension becomes essential as illustrated in Ferrer-i-Carbonell and Frijters (2004).

1.2.1 Property of subjective variables and the use of the panel data

First, what are the properties of subjective variables? The mainstream data sets, such as the BHPS, GSOEP and so on, provide the subjective variables in the form of answers on ordinal scales. Researchers, working in different fields hold different opinions about the use of these variables. For a quantitative evaluation purpose, economists assume that the distance between category i and category j is not necessarily equal to the distance between category j and k so that economists tend to obey the ordinal nature of these variables. Modeling ordinal variables is then based on a non-linear framework using ordered probit or logit models. However, psychologists most of the time neglect the ordinal nature of these variables and proceed by ordinary linear regression. A medium term approach is provided by van Praag and Ferrer-i-Carbonell (2004), the "POLS" (probit OLS) approach. It transforms the ordinal scale into a real scale using the observed frequencies. The transformed variables could thus be treated as cardinal ones with simple linear regressions. This method is extensively used in the studies conducted by the Leyden school researchers, such as the authors of van Praag and Ferrer-i-Carbonell (2004).

The second question is to know if people report their feelings using different scales. In that case, the interpretation and the comparability of these variables become problematic. We have to take into account individual heterogeneity. Panel data provide the solution. Individuals are assumed to be heterogenous, but are supposed to use each a consistent scale over time. The problem of comparability then summarises to the question of individual effects. This is the point made in Ferrer-i-Carbonell and Frijters (2004). However the question is now how to model individual effects when using ordinal variables.

In a probit or logit model, modelling individual effects using fixed-effect approach also has problems because the usual approach spoils a lot of observations (see e.g. Wooldridge 2001). In an ordered probit or logit, the question is even more complex, see Ferrer-i-Carbonell and Frijters (2004). So the usual solution which is adopted with ordered probit model is random effects. However random effects are not without problem as they can be correlated with exogenous variables, leading to estimation biases.

In some of Clark's studies (see e.g. Clark and Georgellis 2013), the ordinal responses are treated as cardinal variables and then a simple regression could be applied and the fixed effect could be identified to cope with the unobserved individual heterogeneity. Similar to the "POLS" solution, the simple cardinalisation greatly reduces the estimation difficulties. Then what is interesting is to explain the differences between the subjective scores reported at two points of time (which corresponds to the fixed effect mechanism) by explanatory variables. Of course, the fixed effect approach has also some limitations. Only time-varying covariates effects are identifiable, some important time-invariant information such as gender, education have to be sacrificed (or condition on them). One solution is proposed in Boyce (2010). In this paper, Boyce uses a three-stage regression to cope with the time-invariant variables in panel fixed effect estimation. In the first step we run a fixed effect regression. The residuals obtained in the first step are predicted by observed (especially the time-invariant covariates) and enter as an observable variable in the final equation.

Compared to the fixed effect solution, the random effect approach depends on extra assumptions on the distribution of the individual effects, such as normality and orthogonality with the covariates. In order to satisfy the condition that the random effect is orthogonal to the covariates, a Mundlak transformation (REMT, Mundlak 1978) could be introduced, an example of the REMT use in SWB study could be found in Ferrer-i-Carbonell (2005) using the GSOEP data. The random effect approach has some technical advantages. First, it is relatively easier to estimate in non-linear framework, thanks to new algorithms such as simulation methods. Second, we do not meet the trade-off between individual effects and time-invariant covariates as in the fixed effect approach. The disadvantage is also very clear, in order to use random effect, we assume that respondents use the same scale when they report their subjective feelings. This assumption might hold in studies of group of people who come from the same community, e.g. citizens of one country who have similar culture and religion backgrounds. Although responses reported from different individual are interpersonally incomparable, people from one country might still behave in a similar way.

The final question concerns international comparisons. We already alluded to the fact that US citizens value personal effort more while European citizens value circumstances more (Alesina and Glaeser 2004). We can thus arguably suppose that citizen of different countries use different scales for answering SWB questions and that these differences cannot be reduced at simple individual effects. Researchers then look forward for some other information for anchoring scales. One solution is to use the vignette examples to measure the anchoring scale used by respondents, see for instance Hopkins and King (2010) and an application in King et al. (2004). Unfortunately, very few data sets are compatible with this approach.

1.2.2 The evaluation of dynamics

What kind of information we might extract from individual data sets? The human thought is based on experiences and expectations rather than on what individuals perceive instantly. Thus the dynamic of social status is one of the most interesting aspect in subjective studies. There are many different types of dynamics we could extract. For example, the occupation status of respondents and that of their parents, reported in cross-sectional data sets such as the GSS for US, the CGSS for China or in panel data set such as the BHPS for UK, the RLMS for Russia and the GSOEP for Germany. This information help researchers to build the inter-generational mobility, e.g. Clark and d'Angelo (2008a) examines the effect of upward inter-generational mobility upon satisfaction and preference for redistribution. The occupation recorded in the data sets are discrete or continuous, and the corresponding classification are usually built according to some criteria thus to reflect the social ladder. For example, in the Erikson-Goldthorpe-Portocarero (EGP) Schema, the classes are discrete and ranked on the basis of two dimensions: employee monitoring difficulties and human asset specificity. Note that this classification is, if it is widely accepted for Western countries might encounter failures when applied to other societies. For instance Wu (2007) reported a failure example of the EGP classification with a Chinese data set.

Income dynamics are also very important. In panel data, the income variables for each individual are usually repeatedly recorded. This gives access to the measurement of the intra-generational income dynamics. Some researchers measure and predict the income dynamics by standard procedures such as a Mincer equation (Idema and Rueda 2011) or an autoregressive model (the Markov-Galton model of Hart 1976). Idema and Rueda (2011) measure the determinants of the income and then predict the average income for each observation during the rest of the life-cycle. However, this method may not be very suitable for predicting the effect of income dynamics because it ignores some non-linear effects. Another way of measuring the dynamics could be found for instance in Alesina and La Ferrara (2005). The authors investigate the preference for redistribution by using the GSS data set. Since GSS is not a panel, they turn to extract the income dynamics from the US Panel Survey of Income Dynamics (PSID) data set by estimating an unconditional Markov transition matrix. The use of the Markov transition matrix sacrifices some information by transforming the continuous income variable into income categories (income intervals or income quantiles). However, this method would show directly the average mobility of the society with the possibility to evaluate some mobility properties. Combining the two papers of Banabou and Ok (a Benabou and Ok 2001a and Benabou and Ok 2001c), economists were especially interested whether the mobility process is progressive or not which reflects the opportunities and risks of social mobility. Moreover, there are clearly some other ways to estimate social mobility in a more general way. One idea is to use dynamic multinomial logit/probit models with explanatory variables where a Markov transition matrix could be indirectly estimated by introducing the lagged income categories. The dynamic multinomial logit model assumes that the error for each category is independent from the others. It also assumes that the odds for preferring one category over another one do not depend on the presence or absence of other "irrelevant" alternatives (independence and irrelevant alternative assumptions, IIA). This is mathematically a convenient property but rather unrealistic in some circumstances. In income category studies the IIA assumption is not a problem. In a non-linear framework such as a multinomial model, it is very difficult to control for individual effects. Honoré and Kyriazidou (2000) provides one solution to cope with fixed effect via conditional likelihood estimation. This method requires that each valid individual has at least four uninterrupted periods of income observations, which would imply dropping a significant part of the information if the panel is unbalanced. The random effect is easier to be applied within a multinomial probit model. There are several ways to do that, such as a simulator estimation. Introducing explanatory variables allows us to relax the homogeneous mobility assumption. We should also notice that the determinants of the mobility itself is a very interesting micro-econometric problem. Some work has to be done in order to do some improvements in this topic. As first pointed out in Heckman (1981a), the distribution of the initial status is not random which needs to be modelled. Yet, the initial condition problem is not very often considered in dynamic multinomial models.

1.2.3 Endogeneity

As in many empirical studies, we met the endogeneity problem. Endogeneity can arise as a result of measurement error, in dynamic models with autocorrelated errors, or simply as the result of omitted variables. The standard way to cope with the endogeneity problem is to use instrumental variables (IV), if the dependent variable is continuous. The advantage of the IV estimation is that with a proper instrument, we are able to reduce the complexity of the causalities and the IV is easy to apply technically. However, how to find a proper instrument and the validation of the instrumental variable are always not an easy task. The endogeneity problem in non-linear models is even more complex. For instance in the 2SLS model, the error distribution of the dependent variable conditioned on covariates is truncated if the endogenous variable is discrete.

van Praag and Ferrer-i-Carbonell (2004) provide one solution in the study of the correlation between overall satisfaction and domain satisfactions. It is necessary to consider the possibility of endogeneity because the degree of satisfaction in different domains are jointly determined by unobservable psychological traits. The authors then suggest to isolate (via a principal components analysis) the common variable (the unobservable psychological traits) that is responsible for the error correlations.[1] Thus a new variable could be constructed and introduced into the overall satisfaction equation. The most important advantage of this approach is that all steps are simple standard procedures and that it is not at all sensitive to the scale of the problem (as. we have many domain satisfaction here). We shall also be cautious that this method (recursive) is only compatible with single direction causality. Thus to say, the exogenous variables explain the variation of the overall endogenous variable (and not the reverse) while they are also correlated with the the overall equation error. This assumption is not realistic if each dependent variable is also correlated with the others (mutually endogenous, which is

1. The errors could be extracted from each independent domain satisfaction equation via POLS estimation (each dependent variables is linearly transformed).

quite usual in subjective analysis).

For example, in political economy and philosophy, researchers are especially interested in the relationship between the redistribution willingness and the perception of the origins of poverty, especially the role of "efforts" and "circumstances". These variables are indeed mutually correlated. A change in one variable is associated with changes in other variables. Thus it is not possible to evaluate the determinants of each variable independently. This leads to the requirement of estimating an endogenous system simultaneously. The estimation of the simultaneous system could be achieved via different ways. When the dimension of the system is only two (two dependent variables), a bivariate distribution could be measured directly. When the dimension is higher than two, the simulation method is preferred. The simultaneous system is a straightforward method who allows to evaluate the unobservable correlations. This method also meets some limitations. First, the computational burden increases dramatically with the dimension of the system. Since this method decomposes the correlation matrix into combination of independent random variables and then simulates them, we need to impose several constraints on the correlation parameters to make sure that correlation matrix is always positive definite (Cholesky decomposable) otherwise the computation is inefficient. Unfortunately, these constraints are not applied in most of the simulation studies. Second, it is hard to consider the mutual determinant in the simultaneous system that a series incoherency problem might arise (we have to obey some mathematical conditions to make sure that the event probabilities sum to 1, see for instance Hajivassiliou et al. (2011)).

It is clear that the subjective approach is not an application of simple regression methods. As an interdisciplinary domain, the subjective approach is a bridge between different theories (economics, sociology, psychology, politics) and empirical work needing important individual panel data sets.

1.3 The structure of this dissertation

In chapter 2, we shall use some standard econometric tools to discuss the relative income and SWB using the BHPS data set, as well as the identification of opportunity and risk perceived from the relative income theory. Chapter 3 gives a revision of the definition of the International Poverty Line (IPL). By using a Bayesian approach and a small sample, we manage to define an IPL with a much better precision along with some modified definitions of poverty in the least developed countries. Chapter 4 discusses the "POUM" effect in the BHPS data set as perceived by different group of people with different job status. In chapter 5, we turn to use the Chinese General Social Survey (CGSS) to investigate the correlation between preference for redistribution and the perceptions of the origins of poverty. Chapter 6 discusses and reviews a side work which discusses and revise the "two tiered dynamic Tobit models with random effect" model (initially provided in Chang 2011b). In this chapter we will show the consequence of some mis-specifications of the model (censoring rule for dynamic variables) and the choice of a more practical way to cope with the initial condition problem in dynamic Tobit models.

Chapter 2

Reference Groups, Reference Income and Inequality Perception

2.1 Introduction: reference groups

Individual utility functions are conventionally seen as a function of income or consumption and leisure. If we look at the domain of happiness economics (see the surveys of Frey and Stutzer 2002, Clark et al. 2008 among others), the relation between income and the level of reported satisfaction is ambiguous. Empirical studies have found only a weak correlation between income and individual well-being. The main focus is provided by the Easterlin paradox (Easterlin 1974) which has gained a revival in the literature with the major paper of Stevenson and Wolfers (2008). The paradox is stated as follows: At a given point in time and for a given country, richer people are happier than poorer ones, but as time passes an increase in GNP does not correspond to an increase in average happiness. Stevenson and Wolfers (2008) provided a very long study, showing that the paradox came from a lack of good quality data and that finally there was a strong relationship between the log of income and life satisfaction in time series. However, this point of view is challenged for instance in Clark et al. (2014), so that there is still room for alternative explanations as those for instance surveyed in Clark et al. (2008). We choose among them the theory of the reference group. If most individuals react positively to an income increase, they mainly pay attention in the longer term to the position of their income with respect to the mean income inside a reference group to which they think they belong. Preferences become interdependent, which is at odds with the conventional view of individual utility theory. Individual happiness and satisfaction depend on what one

0. This paper was co-authored with Michel Lubrano.

achieves in terms of a comparison with others (Ferrer-i-Carbonell 2005). A higher status entails positive effects for subjective well-being while a relative low status entails negative effects.

Reference groups are have become a major topic in the happiness literature. Using the comparison theory, economists and psychologists have tried to explain the Easterlin paradox in empirical studies, see e.g. Blanchflower and Oswald (2004), Clark and Oswald (1996), Easterlin (1974), Ferrer-i-Carbonell (2005) or Frey and Stutzer (2002). This gives us a good reason to investigate what constitutes a reference group, what is its definition and contents and what are the possible conclusions. An essential question that might have been ignored in the study of reference groups, (see however the recent paper by Clark and Senik (2011)), is the sensitivity of the results to the definition of the reference group, i.e. to which groups do people compare themselves? Does the comparison target maintains stable in different situations and periods?

We have several aims in this paper. We first want to review the existing various possibilities for defining first a reference group and second a reference income in order to measure their incidence on empirical results. Second, most if not all empirical studies report an elasticity of compensation income which is much higher than 1. This means that for instance if the reference income is increased by 10%, personal income has to increase by far more than 10% in order to keep the same level of well-being. This result is counter-intuitive, except if the reference income measures something else than just a monetary reference. Third, and more importantly, what we shall show is that not only is the reference income important, but also its dispersion within-group. A reference group is a complex object containing a lot of heterogeneity.

The paper is structured as follow. Section 2.2 briefly discusses the definition of comparison income and some relating models based on “absolute” versus “relative” income. Section 2.3 introduces the framework of subjective well-being data and the econometrics treatment. Section 2.4 presents the data and the basic estimation of well-being following with variant models focusing on the asymmetric effects due to individual heterogeneity. Section 2.5 intends to analyse inequality effects within two reference groups. Section 2.6 concludes.

2.2 A survey of comparison income definitions

A reference group is a collection of individuals or households that share some common characteristics which are either objective or subjective. The common characteristics can be a similar level of income, belonging to the same place of employment, to the same neighbourhood, region or country (see e.g. Clark et al. 2008 for a discussion). Let us assume for the moment that the reference income $y^r = \bar{y}$ is simply the within group average income in order to discuss how the reference income can enter the individual utility function. We want to formalise the idea that income enters the utility function in two different ways: current income y_i and comparison income as the ratio y_i/y_i^r. We have a similar formulation for instance in Clark et al. (2008). Using a panel notation, we have:

$$u_{it} = \beta_1 \log(y_{it}) + \beta_2 \log(y_{it}/y_{it}^r) + \beta_3 x_{it} + \epsilon_{it}. \tag{2.1}$$

In this equation, y_{it} stands for the current income of individual or household i at time t, x_{it} is a vector of personal characteristic variables and y_{it}^r is the reference income while u_{it} is of course the unobserved utility level. Over time, economic growth increases the level of both the individual income and of the reference income. As a consequence, an individual benefits from economic growth if and only if $(\beta_1 + \beta_2)\Delta \log(y_{it}) > \beta_2 \Delta \log(y_{it}^r)$. Recalling the findings of Osberg and Sharpe (2002), in some developed countries the inequality is increasing and the increase in personal income is limited to the upper tail of the income distribution, see e.g. the UK and the US. This increasing inequality will negatively affect most people; those who have the lowest income increase will lose some of their well-being due to their declining relative status. This might be a complementary explanation to the Easterling paradox. A part of the increase in total income is wasted for well-being because of the asymmetry in the income distribution. The above equation does not enable this type of explanation because only mean level of income of reference group is myopic. We shall propose a solution in section 2.3.

2.2.1 Subjective reference groups

As we are in a context where well-being is self-reported, it would be natural to ask individuals to report also what they consider to be their own reference group. There exist very few studies using this approach, mainly because large public panels do not incorporate such a question. We can note however that Melenberg (1992) used the Dutch Socio-Economic Panel where individuals are asked in 1985 and 1986 to define the "people whom you meet frequently, like friends, neighbours, acquaintances or possibly people you meet at work". These data are now quite dated. A more recent paper is Knight et al. (2009) which uses a Chinese survey that contains the following question: "generally speaking, to whom do you think you compare yourself to mostly?". In this survey launched at the end of the 1990s, 68% of the respondents reported that their main reference group consisted of individuals living in the same city. The two more important panel surveys in Europe, the British BHPS and the German GSOEP (except for some rare periods for the latter) do not include such an information. This limits very much the usefulness of this approach. If we limit our attention to cross section data, in one wave of the European Social Survey contains questions about reference groups and also about the intensity of comparison. Exploiting these data, Clark and Senik (2011) found, among other things, that colleagues at work are the most frequently cited reference group and that well-being decreases with the intensity of the comparison.

2.2.2 Researcher defined reference groups

The other branch of the literature considers as a reference income the income of "people like me". This is the most frequently used method. One needs first to define the reference group before estimating a work or life satisfaction equation. This is the "researcher defined" reference group approach. In this framework, the reference income can be calculated in two different ways:

- We can estimate a general Mincer wage equation and then compute the predicted wages of "someone like me" (see e.g. Clark and Oswald 1996). This means comparing individuals having the same human capital (education and experience).
- We can define cells by considering individuals having the same broad characteristics such as age, education level, gender or living in the same region (East and West Germany for instance). Once the cells are determined, the reference income is defined as a central tendency for each group, usually the mean, but why not the median. This will be the method used in this paper. See also Ferrer-i-Carbonell (2005) or Cappelli and Sherer (1988).

We must however note that several other rationales could be used for this selection. For instance, at an aggregate level, Peng et al. (1997) noted that people from different cultural groups use different referents in their self-reported values. E.g. Chinese people compare to other Chinese whereas Americans compare to other Americans. At an individual level in UK, Clark (1996) relates answers to a job satisfaction question with wages of partners and to average wages of other household members. McBride (2001) introduces a family reference income, using the question contained in the GSS referring the income of the parents in order to characterise social mobility between generations.

2.2.3 Characterising the reference group

In most papers, the variables which are used to define "people like me" are not discussed with respect to a particular economic theory. For instance, Ferrer-i-Carbonell (2005) uses education, age, but also region and eventually she tested the significance of gender. So the precise definition of the reference group is not seen as important. However, the estimation results of an equation like (2.1) can be sensitive to this definition. In most data sets like for instance the BHPS, most of the sampled individuals have an income which mainly comes from earnings and marginally from social benefits. The presence of capital income is very scarce. Consequently, “the people like me” can be supposed to be the people that have the same human capital. In this case, the average cell income would represent the average earnings that corresponds to the average human capital. We are not therefore very far from a Mincer equation. This has the consequence that the main variable defining a group is the education level.

2.3 Economic and econometric assumptions

Ordered probit models are designed for analysing answers to a question where the possible items are ordered and discrete. Econometricians have promoted the use of this model for analysing survey data while psychologists have a tendency to prefer ordinary least squares models which require an implicit cardinality assumption. These models have been extended to deal with panel data, the main goal being to cope with individual effects. Individuals with the same characteristics may not answer questions in the same way. However, when using panel data, we have also access to another dimension which is income dynamics. In order to relate well-being answers to observed characteristics including income dynamics, a certain number of economic and econometric assumptions have to be made that we shall now set out. See Ferrer-i-Carbonell and Frijters (2004) for an view.

2.3.1 Basic model

Let us consider a set of individuals who are reporting life satisfaction levels noted W_i. These levels are at value on a Cantril scale, which means that these levels are ordered and that the scale is represented by numbers between for instance 1 and 7 (BHPS) or 0 and 10 (GSOEP). For the BHPS, the question is: *Using the same scale, how dissatisfied or satisfied are you with your life overall?* On this scale, 1 corresponds to *Not satisfied at all* while 7 corresponds to *Completely satisfied*. The anchoring of the scale is left to responder. A life satisfaction question can be phrased differently as reported for instance in Helliwell and Wang (2012). The different items are explicitly given and can be for instance: *fully satisfied, fairly satisfied, just satisfied, not very satisfied, not at all satisfied*. These items are then recoded on an ordered numerical scale. Finally, according to Larsen et al. (1985), a happiness question (how happy you are) give less reliable answers than a life satisfaction question.

In order to devise a relationship between reported well-being W_i and utility u_i, we have first to assume individual consistency:

A1 *The reported levels W_i are related to the unobserved levels of welfare or utility u_i in a consistent way which implies that if the W_i for a given individual changes over time, this change is consistant with an individual change over time in u_i.*

As we are observing different individuals in the same sample at a point of time, we have to be able to assume at least ordinal comparability between them, which requires a further assumption:

A2 *Individuals use a common evaluation scale, so that for two individuals i and j*

$$W_i > W_j \Rightarrow u_i > u_j \qquad \text{for } i \neq j,$$

implying ordinal comparability.

For detailed psychological discussions of this assumption, see Sandvik et al. (1993), Diener et al. (2003). With these two assumptions, we can accumulate statistical information.

If we want to implement these two assumptions (consistency and ordinal comparability), how can we use the reported levels W_i in order to infer utility levels and their relation to a set of personal variables? The econometric literature has proposed the ordered probit model which, for K categories estimates $K-1$ unknown levels μ_k such that:

A3 *The W_i are first related to the unobserved utility levels using a set of inequalities*

$$\begin{array}{ll} W_i = 1 & \text{if } u_i < \mu_1 \\ W_i = 2 & \text{if } \mu_1 < u_i < \mu_2 \\ \cdots & \\ W_i = K & \text{if } u_i > \mu_{K-1}, \end{array}$$

The unobserved utility levels u_i are then explained by a set of observed personal characteristics:

$$u_i = x_i\beta + \epsilon_i, \tag{2.2}$$

where the ϵ_i are assumed to be normal distributed with zero mean and variance σ^2.

The normality assumption can be relaxed as in e.g. Stewart (2004). Assumption A2 can be relaxed with the use of panel data.

2.3.2 Panel data models

Panel data do bring in a new dimension. We observe the same individuals over time which allows us to relax slightly the assumption of interpersonal comparability as we can allow for individual heterogeneity. For instance, some individuals are optimistic while some others are pessimistic. This means that they can report a different level of well-being while having the same socio-economic characteristics. The only maintained assumption is time consistency:

A4 *Individuals with the same characteristics can have different well-being evaluations, using an evaluation scale which has only to be time independent. Individual effects v_i are introduced in the regression equation:*

$$u_{it} = x_{it}\beta + v_i + \epsilon_{it}. \tag{2.3}$$

in order to take into account unobserved individual heterogeneity.

Time consistency means that being optimistic does not depend on age. We note that for the while individual effects are additive, they modify only the constant term of the regression, or alternatively the unknown thresholds μ_k. Ferrer-i-Carbonell and Frijters (2004) found that it was more important to take into account individual heterogeneity than the discrete and ordinal characteristics of the data. [1]

2.3.3 Panel data and income dynamics

The v_i individual effects can be either fixed or random. Following Rendon (2012), the sole difference between the two options is prior information. With a random effect, we suppose that the v_i are constrained by having a common $(0, \sigma^2_{v_i})$ Gaussian distribution while with a fixed effect model, the v_i are independent constants and specific to the sample being used. In the case of random effects, the crucial assumption is that both the ϵ_{it} and the v_i are

1. We can also introduce a time fixed effect common to everybody indicating to which period each observation belongs. Each year can have specific characteristics such as different macroeconomic shocks, but more simply the time effects are a simple way to take into account inflation. This is done by introducing αT_t (where T is a matrix of zero and ones with as many columns as there are periods in the panel).

independent of the x_{it}. This assumption is logical for the ϵ_{it}. It is however too strong to suppose that the individual effects v_i are independent of all the individual characteristics such as income. We can however suppose that the v_i are independent of the age or the gender of the individuals. A conventional solution is to model the correlation between a smaller subset of the mean value of x_{it} over the time dimension and the v_i. We are going to suppose that the subset of x_{it} is just income, y_{it}, leading to the following assumption:

A5 *Individual effects are correlated with long term personal income and are independent of the other individual characteristics.*

The correlation between income and the individual effect is modelled by:

$$v_i = \overline{\log y_i}\lambda + \eta_i,$$

where $\overline{\log y_i}$ is the mean over t of $\log y_i$ and the η_i are now supposed to be uncorrelated with the other explanatory variables. This is the solution advocated in Mundlak (1978) and used for instance in Ferrer-i-Carbonell (2005). The original model is transformed into:

$$u_{it} = x_{it}\beta + \overline{\log y_i}\lambda + \eta_i + \epsilon_{it}. \tag{2.4}$$

The term $\overline{\log y_i}$ is considered as a simple statistical correction term.

However, we would prefer to have a model where each variable has a clear economic interpretation. One of the possible many explanations to the Easterlin paradox is that individuals do not react to the level of their income (which is usually I(1) when the answers to the satisfaction question are obviously I(0) because they are on a bounded scale), but to the variation of their income, $\Delta \log y_{it}$. When y_{it} is replaced by $\Delta \log y_{it}$, we have a balanced relation as now both $\Delta \log y_{it}$ and W_{it} are integrated of order zero. This explanation is a complement to the reference income explanation. We just have to transform the current income variable into the sum of a transitory variation, $\Delta \log y_{it}$ and of a long term or permanent income which is $\log \bar{y}_i$ and no longer $\overline{\log y_i}$ as in the Mundlak correction if we compute the permanent income as the mean income over the period. We have gained a solid economic interpretation. So instead of A5, we propose A6:

A6 *Individual utility depends on income through the short term variation of income, the long term permanent income and the reference income with:*

$$u_{it} = \beta_1 \Delta \log y_{it} + \beta_2 \log \bar{y}_i + \beta_3 \log y^r_{it} + \gamma x_{it} + \eta_i + \epsilon_{it}. \tag{2.5}$$

In this equation, the relative income ratio has to compare the long term individual income $\bar{y}_i$ with the reference income y^r_{it}.

A7 *In a dynamic setting, the long term personal income is compared to the reference income defined as the mean income of the reference group:*

$$u_{it} = \beta_1 \Delta \log y_{it} + \beta_2' \log \bar{y}_i + \beta_3' \log \bar{y}_i / y_{it}^r + \gamma x_{it} + \eta_i + \epsilon_{it}. \qquad (2.6)$$

where $\beta_2' + \beta_3' = \beta 2$ and $-\beta_3' = \beta_3$. The reference income can be defined either as the mean or median income of the reference group. There is a unique reference income for all the individuals belonging to a given group, but this reference income can evolve over time.

The final question is the meaning of β_2' in this equation. If it is positive, we have an income anchoring effect. Economic growth benefits to everybody. A value of zero is the most plausible solution as it means that if the ratio between long term income and reference income remains unchanged, the utility level remains constant, validating the Easterlin paradox. A negative value is certainly an indication of misspecification.

2.3.4 Reference income and income inequality

The only comparison term in (2.6) is the distance between the long term personal income and the reference income. The shape of the income distribution either inside the reference group or as a whole is not taken into account. In many countries, the increase in personal income was limited to the upper part of the income distribution. Those who are at the lower part of the income distribution will lose some of their well-being due to their declining relative status. If the reference group is defined according to education and if the increase in income is limited to the highest educated individuals, we might well discard this effect by just looking inside each reference group and ignoring what happens between the groups.

Before discussing the way to introduce a measure of inequality in our well-being equation, we must go back to the fundamental question of the representation and meaning of inequalities which was first raised by Rawls (1971). An inequality can be felt as just if it rewards effort and talent. In this case, inequality represents an opportunity. If in the education same, individuals can expect different wages depending on their effort, we can suppose that these expectations make them happier. On the other hand, inequality is felt as unjust if it concerns factors for which individuals are not responsible such as for instance handicap, social origin and so on. In this case, inequality is a risk for which individuals have to be compensated by society. In particular, inequality resulting from discrimination and lack of capacities is felt as unjust following Sen (1993). The empirical question is then to disentangle these two types of inequalities, to find an identification rule.

The empirical literature is rich in contradictory results, see Senik (2005) for a survey, due to the by lack of such an identification rule. Measuring inequality for the whole population with a Gini index would produce a single number that could not be disentangled from the constant term. In order to introduce variability, we have to measure inequality within a predefined group. If a reference group is defined by education, individuals freely chose to belong to that group when they decide to educate. The reference income in this case represents the average reward to a given stock of human capital and inequality represents opportunities of a future reward based on effort. If a reference group is defined independently of education, choosing regions for instance, then we can assume that individuals are distributed at random within those regions and groups, at least if they do not move. Those groups will contain a mix of different education levels and of different incomes. Consequently inequality within these groups can be assumed to represent overall inequalities that are generated by other factors than individual decisions. We can then suppose that inequality measured within those groups can identify inequality as a risk.

A8 *Individuals have different reference groups from which it is possible to identify different attitudes to inequality:*

$$\begin{aligned} u_{it} &= \beta_1 \Delta \log y_{it} + \beta_2 \log \bar{y}_i + \beta_3 \log \bar{y}_i / y^r_{it} \\ &\quad + \beta_4 Gini^r_{it} + \beta_5 Gini^{r'}_{it} + \gamma x_{it} + \eta_i + \epsilon_{it}, \end{aligned} \tag{2.7}$$

where $Gini^r_{it}$ is a Gini coefficient computed within the first reference group used to compute the reference income while $Gini^{r'}_{it}$ is a Gini coefficient computed within a second reference group, independent of the first one.

2.3.5 Identification and likelihood function

The likelihood function of the simple ordered probit model is based on the normality assumption for the ϵ_{it} from which we compute

$$\begin{aligned} \text{Prob}(W_i = k) &= \text{Prob}[\mu_{k-1} < x_i\beta + \epsilon_i < \mu_k] \\ &= \text{Prob}[\mu_{k-1} - x_i\beta < \epsilon_i < \mu_k - x_i\beta] \\ &= \Phi(\frac{\mu_k - x_i\beta}{\sigma}) - \Phi(\frac{\mu_{k-1} - x_i\beta}{\sigma}), \end{aligned}$$

where $\Phi(.)$ is the Gaussian cumulative distribution. The likelihood function is written as

$$\log \mathbf{L} = \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{1}(W_i = k) \log[\Phi_{ik} - \Phi_{i,k-1}],$$

where $\mathbb{I}(.)$ is the indicator function. Maximisation of this log-likelihood function cannot lead to a unique solution without additional identification restrictions. Without any constraints on β, μ or σ^2, the outcome of log-likelihood maximisation would endlessly circle on a plateau of equally-likely combinations of β, μ or σ^2. Identification can be obtained in different ways. A first constraint is given by imposing $\sigma^2 = 1$ as $\Phi(\frac{\mu_k - x_i\beta}{\sigma}) - \Phi(\frac{\mu_{k-1} - x_i\beta}{\sigma})$ has only one set of estimates of β and σ that maximize the likelihood. A second set of constraints has to be imposed on the threshold. We cannot have at the same time free thresholds parameters and a free constant term in the regression. So, in general, we impose the a zero restriction on the first threshold parameter μ_1. But excluding a contant term from the regression is an alternative possibility. With these identification restrictions, we can obtain the MLEs of β and of the thresholds μ_k.

The panel dimension introduces some complications which comes mainly from the random individual effects:

$$\begin{aligned}\text{Prob}(W_{it} = k) &= \text{Prob}[\mu_{k-1} < x_{it}\beta + \eta_i + \epsilon_{it} < \mu_k] \\ &= \Phi(\mu_k - x_{it}\beta - \eta_i) - \Phi(\mu_{k-1} - x_{it}\beta - \eta_i).\end{aligned}$$

The contribution of one individual to the likelihood function is given by

$$\int \phi(\eta_i|0,\sigma_v^2) \prod_{t=1}^{T} \prod_{i=1}^{N} \left[\Phi(\mu_k - x_{it}\beta - \eta_i) - \Phi(\mu_{k-1} - x_{it}\beta - \eta_i)\right] d\eta_i,$$

where $\phi(\eta_i|0,\sigma_v^2)$ is the distribution of the individual effects. This equation involves the computation of a one dimensional integral. According to Butler and Moffitt (1982), there are simple ways of computing this integral; see also Crouchley (1995) for a general treatment. As long as the dynamics are confined to the income variable, there is no additional problem of estimation.

2.4 An investigation using the BHPS

The British Household Panel Survey (BHPS) provides a sample of more than 6000 British households first interviewed in 1991. The members of these original households have since been followed and annually interviewed. We extracted a balanced panel covering the years 2002-2008 and corresponding to 3 311 individuals (the satisfaction question is initially inserted in BHPS since the year 2002). We want to address several empirical questions in this paper. We want first to see if a good specification of income dynamics can explain a part of the Easterlin paradox and what is its relative weight compared to the reference group explanation. Second, we want to explore the sensitivity of the results to the specification of the reference group. Third, the effect of the reference group is certainly non-linear and various specification for non-linearity have to be tested. Finally and most importantly, rather than simply introducing a single characteristic of the reference group (normally measured by mean or median income of the reference group), we enquire whether subjective well-being responds to other possible characteristics of a reference group, and in particular to the dispersion of income within the reference group and if the impact of overall inequality can be measured.

2.4.1 Income dynamics

We now introduce a simple model of life satisfaction including income dynamics, but not including for the while a reference income. Using Equation (2.5) where we have dropped $\log y^r_{it}$, we get our starting equation with estimation results presented in Table 2.1. Time dummy variables are significant even after deflating income for inflation.[2] Age enters in a non-linear way, producing a U-shape which means that well being decreases till the age of 40 and increases after that age. This is in accordance with the results found in Blanchflower and Oswald (2008). The income variables enter the equation with the correct positive sign, but are not very significant. Transitory income variations have a lower impact than permanent income. But both coefficients are rather small. The permanent income is measured by the mean log absolute income of an individual over t, and is denoted as $\bar{y}_i$. It enters the equation with a positive coefficient 0.061. The transitory income $\Delta \log(y_i)$ has a positive coefficient 0.046. So total income effect is $0.061+0.046 = 0.107$. Thus life satisfaction depends mainly on age and health status, on family composition and only marginally on income dynamics.

2.4.2 The choice of a reference group definition

We are going to introduce reference groups and reference income in order to estimate our full model (2.5). In this estimation, we will define the reference group on *a priori* grounds (researcher defined). The goal is to measure the influence of the comparison income on life-satisfaction. We have argued in section 2.2 that we should define a reference group with respect to human

2. Household incomes were adjusted by the following price index: 2002, 95.4; 2003, 96.7; 2004, 98; 2005, 100; 2006, 102.3; 2007,104.7; 2008, 108.5. Ferrer-i-Carbonell (2005), for a similar empirical question, has used the German panel GSOEP for studying the effect of reference income on subjective well being with fixed reference groups (and presumably an unbalanced panel). She advocate the use of time dummies as a substitute to price deflators.

Table 2.1 – Estimation of a first life satisfaction equation

	Estimate	t value
Intercept	20.152	9.101
date2004	-0.031	-1.200
date2005	-0.133	-5.053
date2006	-0.068	-2.581
date2007	-0.056	-2.073
date2008	-0.042	-1.543
log(age)	-9.276	-7.561
log(age)^2	1.257	7.418
Min age	40.0	
marriage	0.487	13.377
log(adults)	-0.206	-5.845
log(1+kids)	-0.082	-2.699
health	-0.388	-29.832
$\Delta \log(y)$	0.046	1.925
$\log(\bar{y})$	0.060	1.513
μ_1	0.585	15.562
μ_2	1.262	30.043
μ_3	1.987	45.747
μ_4	3.046	68.250
μ_5	4.452	94.435
σ	1.105	54.024
Log-likelihood	-25011.71	
N	3311×6	

capital characteristics. Let us start with education categories [3] and continue with age brackets to take into account the life cycle. [4] Gender can be a further variable to consider. As we are using a panel, some variables defining the reference groups change over time, such as age and marginally education while gender remains constant. We shall experiment with 4 different definitions of the reference group:

1. Model 1: Education and waves: 9 education categories and 6 periods, we have 54 different cells.
2. Model 2: Education, gender and waves: 9 education categories, 2 genders and 6 periods leads to 108 cells.

3. The education level is classified as $1a, 1b, 1c, 1a, 2b$, $2c\_gen, 2c\_voc, 3a, 3b$ following the CASMIN educational classification. For more details see appendix A.1.1.

4. Age brackets are: 16-20, 21-30, 31-40, 41-50, 51-60 and over 61 years old.

3. Model 3: Education, waves and age brackets: 9 education categories, 6 periods and 6 age brackets leads to 324 cells.
4. Model 4: Education, gender, waves and age brackets. 9 education categories, 2 genders, 6 periods and 6 age brackets leads to 648 cells.
5. Model 5: Region and waves: 19 region regroups and 6 periods, we have 114 different cells.

In Model 1, we assume that individuals compare their income only with individuals belonging to same education category, with possible changes over time. People inside the same reference group are supposed to have equal opportunities or capacities. With Model 2, we assume that men and women can have different opportunities. Males compare to males and females to females. With Model 3, we take into account their life-cycle, but not gender differences. Individuals do not have the same expectations at different points of their life cycle. They compare themselves, in term of opportunities to individuals of the same age group. Model 4 considers a complete specification with education, life cycle and gender.

In the literature, the comparison income is always taken as the mean of the reference group so that it is sometimes called the mean reference income. However, it is very easy to find that the income distribution within every group can be very asymmetric. So it could make a difference to compute the mean or to compute the median. The median is in a way more representative of a centrality indicator as it does not depend on extreme values.

The sample size is $3311 \times 6 = 19\ 866$ which makes on average between 368 individuals per cell for the simplest model and 31 individuals per cell for model 4. We report in Table 2.2 the estimation of the three income variable coefficients. The reference income always appears with a negative sign, as expected while the two other coefficients remain positive. We have checked,

Table 2.2 – Four models of life-satisfaction using median income of different reference groups

		Model 1	Model 2	Model3	Model 4	Model 5
	estimate	0.051	0.051	0.050	0.051	0.048
$\Delta \log(y)$	t-ratio	2.105	2.119	2.075	2.114	2.014
	estimate	0.132	0.135	0.133	0.129	0.091
$\log(\bar{y})$	t-ratio	3.107	3.185	3.186	3.319	2.224
	estimate	-0.420	-0.429	-0.368	-0.329	-0.543
$\log(y^r)$	t-ratio	-3.829	-4.095	-4.526	-4.599	-3.713
log-likelihood		-24999.36	-24997.12	-24999.93	-24999.22	-25003.31

using a Wald test (see Appendix A.1.3), that the five different reference groups did not lead to significant different results at the 5% level. This was true either for the complete regression or just for the three income variable coefficients. Considering the likelihood value, there does not seem either to be a significant difference between the different models. Model 1 is sufficient and other models do not introduce supplementary information on the regression coefficients. [5] Consequently, it is sufficient to consider education levels to define a reference group for comparing incomes.

A striking fact in Table 2.2 is that when we introduce the reference income, the two other income variables become very significant. So we cannot have a separate explanation of the Easterlin paradox using income dynamics with Δy_{it} on one side and on the other side using the reference income. This has to be a joint explanation.

2.4.3 The empirical content of reference groups

Let us have a deeper look at the content of the reference groups defined by education levels. It is often argued that income inequality has remained relatively stable over the period covered by the BHPS when it had experienced large changes in the previous period (see e.g. Jenkins 2000). And also that the last income decile has increased much more than the lower deciles, in the US and also in the UK. Over the period covered by the BHPS, the largest mean income concerns the *high tertiary* category and the mean income of that group has regularly increased. At the other extreme, the *no education* category has the lowest mean income which remained relatively stable.

In order to characterize income dispersion inside each group, we computed a Gini coefficient for each category, grouping this time all the years together. In Table 2.3, the greatest inequality is found in the lowest group (the one with

Table 2.3 – Gini for educational categories 2002-2008

Casmin*	1	2	3	4	5	6	7	8	9
Gini	0.278	0.256	0.243	0.264	0.241	0.269	0.240	0.256	0.253
Rank	1	4	7	3	8	2	9	5	6

*: Casmin education classification, see Appendix A.1.1

no education), followed by high general, middle general groups. Vocational education, whatever its level experiences the lowest inequality.

5. Ferrer-i-Carbonell (2005) finds similar results on German data. She defines the reference group by education, age and region. In an appendix, she shows that, at least for Germany, including gender in the definition of the reference group is not statistically significant.

2.4.4 The puzzle of personal versus reference income

Now that we have chosen the definition of the reference group and reference income, we give the estimation of our full model (2.6) in Table 2.4. We have a puzzle with this version of the model. We would expect that the

Table 2.4 – The full puzzling model

	Estimate	t value
Intercept	22.607	9.707
date2004	-0.020	-0.791
date2005	-0.117	-4.368
date2006	-0.043	-1.572
date2007	-0.021	-0.751
date2008	-0.001	-0.040
log(age)	-9.004	-7.219
log(age)^2	1.214	7.040
Min age	40.8	
marriage	0.478	13.190
log(adults)	-0.250	-6.951
log(1+kids)	-0.078	-2.573
health	-0.393	-30.117
$\Delta \log(y)$	0.050	2.105
$\log(\bar{y})$	-0.290	-2.803
$\log(\bar{y}/y^r)$	0.422	3.857
μ_1	0.585	15.565
μ_2	1.263	30.078
μ_3	1.989	45.794
μ_4	3.048	68.293
μ_5	4.459	94.519
σ	1.104	54.172
Log-likelihood	-24996.17	
N	3311×6	

coefficient of $\log \bar{y}_i$ to be zero or eventually positive once we introduce the relative income (as for instance in Blanchflower and Oswald (2004) for the US). The same increase of the reference income and of the increase of permanent income should be neutral. This means that in equation (2.5), β_2 and β_3 should be equal in absolute value. Obviously this restriction does not hold as $\log(\bar{y})$ has a negative and significant coefficient. β_3 is much larger in absolute value than β_2.

This puzzle might be due to the BHPS data set. Using the GSOEP, van

Praag and Ferrer-i-Carbonell (2004, chap. 8) do report a ratio $-\beta_3/\beta_2$ equal to 1 with reference groups defined by education, age and region. Ferrer-i-Carbonell (2005) reported a similar value using the Westerner subpopulation from the German GSOEP with reference groups defined similarly. Using the BHPS, Clark (2003) found implicitly a value of 5.65 for this ratio while we have here a value of 3.18. A ratio greater than 1 means that we must have a much larger increase of the permanent income in order to keep the same level of life satisfaction. This empirical puzzle suggests that there is a neglected factor in our model when taking into account reference income. We can look in two directions: the presence of nonlinearities in the role played by the reference income (for example being below or above the reference income). Some of these possible non-linearities have already been explored in the literature (see for instance Ferrer-i-Carbonell 2005). The second possibility that we want to investigate concerns the characterisation of the reference group. For now, we have considered only a central tendency indicator with the reference income. But the dispersion of income inside the reference group can play an important role and also present some asymmetries.

2.4.5 Asymmetric effects

Ferrer-i-Carbonell (2005) has detected some asymmetric effects using the GSOEP. She found that for individuals below the mean of their reference group, the β_3 as defined in (2.6) was larger in absolute value than the β_3 corresponding to individuals above their reference income.

Using the BHPS, the answer is not so clear. If we simply modify model (2.6) so as to allow for different coefficients for the income variables depending on whether an individual's income is below or above his reference income while keeping all the other coefficients equal, we do not find the presence of asymmetry. We have to run two completely separate regressions for two sub-populations. Results for an asymmetric model (2.6) are reported in Table 2.5. A Wald test of equality for whole set of coefficients shows significant differences between the two regressions (P-value=0.011). Regarding the magnitude of β_3 between richer and poorer populations, β_3 for poorer is higher ($0.443 > 0.387$) although such a difference is not significant according to a t-statistic (see Appendix A.1.3). But we could say that comparing these two coefficients is not meaningful as we could not impose a unit elasticity ($\beta_2 \neq 0$). So we have to compare the ratio (the permanent income elasticity) $\partial \log(\bar{y}_i)/\partial \log(y^r_{it}) = \beta_3/(\beta_3 + \beta_2)$. In this case, we found 4.55 for the poorer group and 2.68 for the richer group so that the previous comparison is amplified. However, the difference is still not significant according to a t-test (see Appendix A.1.3).

Table 2.5 – Estimation of an asymmetric satisfaction equation

	Below the reference income		Above the reference income	
	Estimate	t value	Estimate	t value
Intercept	22.175	6.838	22.562	6.858
date2004	-0.029	-0.787	-0.011	-0.279
date2005	-0.057	-1.478	-0.180	-4.587
date2006	-0.043	-1.095	-0.045	-1.103
date2007	-0.006	-0.139	-0.040	-0.953
date2008	0.015	0.347	-0.021	-0.470
log(age)	-8.633	-5.039	-9.025	-5.048
$\log(\text{age})^2$	1.173	4.958	1.206	4.881
Min age	39.6		42.0	
marriage	0.475	9.275	0.446	7.129
log(adults)	-0.206	-3.733	-0.336	-5.884
log(1+kids)	-0.062	-1.454	-0.107	-2.497
health	-0.412	-22.583	-0.375	-19.714
$\Delta\log(y)$	0.056	1.650	0.021	0.505
$\log(\bar{y})$	-0.346	-2.506	-0.243	-1.660
$\log(\bar{y}/y^r)$	0.443	3.075	0.387	2.393
μ_1	0.545	11.842	0.662	10.029
μ_2	1.230	23.354	1.335	18.621
μ_3	1.982	36.146	2.034	27.659
μ_4	3.018	52.970	3.115	41.487
μ_5	4.378	71.433	4.572	58.309
σ	1.189	38.061	1.234	36.410
Log-likelihood	-13003.69		-12406.08	
N	9919		9947	

In fact, the main difference between the two regressions in Table 2.5 comes from the thresholds (p-value=0.0014 for a Wald test). That means that individuals in the two groups use a different evaluation scale.

We have formulated our model in terms of relative income ratio with (2.6). The restriction $\beta_2 = 0$ should be imposed, but it is never accepted. Taking into account a first type of non-linearities does not solve our empirical puzzle. We shall now try to complement the reference income by an indicator of inequality.

2.5 The impact of inequality

In a usual welfare function like that of Atkinson (1970), the social planner is supposed to be averse to inequality. In the global development index of Osberg and Sharpe (2002), income inequality enters the formula as a negative factor. And Thurow (1971) argues that "The distribution of income itself may be an argument in an individual's utility function". So there are large incentives to investigate empirically the influence of income inequality on well-being, see the survey by Senik (2005).

Empirical findings concerning the impact of inequality on well-being are mixed. Using the GSOEP (waves 1985-1998), Schwarze and Harpfer (2007) found that a Gini index calculated for the 75 regions of West Germany is negatively correlated with life-satisfaction. Alesina et al. (2004) undertook an international comparison between the USA and Europe. They found that the life-satisfaction of Americans does not respond significantly to inequality using the General Social Survey, 1972-1994. On the other hand, Europeans' satisfaction is found to be decreasing with inequality, particularly for poor and left-wing people, using the Euro-Barometer Survey, 1975-1991. Blanchflower and Oswald (2003) reports similar results. The differences in inequality responses are, according to Alesina et al. (2004): "*...in the US, the poor see inequality as a ladder that may be climbed, while in Europe the poor see that ladder as a difficulty to ascend*". In other words, income inequality can be seen either as an opportunity or as a nuisance, depending on the country. How an individual responds to it depends on culture, status, political ideas, religion, etc. However, these studies fail to identify the possibility of having the two possibilities: inequality as a risk or inequality as an opportunity, depending on how inequality is measured.

2.5.1 Inequality and reference groups

For the UK, we have the result found in Clark (2003) that individuals react positively to inequality when the latter is measured within reference groups. Clark (2003) defined his reference groups with respect to region, gender and waves, which is in a way not so different as what is found in Schwarze and Harpfer (2007) who used regions and waves for defining their groups. So we could have expected a negative sign using the UK data. There is obviously a lack of identification.

As we have defined reference groups with respect to education levels and waves, a positive coefficient for a Gini index can be interpreted as a measure of opportunity for a given education level. Let us introduce a Gini coefficient in our basic equation 2.8 as

$$u_{it} = \beta_1 \Delta \log(y_{it}) + \beta_2' \log \bar{y}_i + \beta_3' \log \frac{\bar{y}_i}{y_{it}^r} + \beta_4 Gini_{i,t}^r + \gamma x_{it} + \eta_i + \epsilon_{it}, \quad (2.8)$$

where $Gini_{i,t}^r$ is the Gini coefficient computed within the reference group of individual i at time t. The results reported in Table 2.6 first confirm that there is ample room for a second indicator characterising a reference group. The reference income, which is a centrality indicator, is still significant and keeps its negative sign with $-\beta_3' = 0.394$. The reference Gini, which is also an indicator of dispersion, appears significantly. So both indicators are needed. Secondly, the Gini coefficient appears with a positive sign (and a value of $\hat{\beta}_4 \approx 2$), confirming that inequality within the educational group can be seen as an opportunity. However, introducing a reference Gini has not yet solved our empirical puzzle as β_2' is still negative and significant. Could a finer specification, allowing in particular for asymmetries, solve our puzzle? In particular, we think that different education groups can react differently to within group inequality. We have seen that the group with no education degree experienced the largest inequality index. Among the low educated individuals (categories $1a, 1b, 1c$), it is the largest group (see Appendix A.1). Table 2.7 show us that the lowest educated group has a different vision of inequality. The impact of the Gini is 1.750 for all the categories while it is equal to 1.750+0.658=2.41 for the lowest educated individuals. We can conclude that low educated individuals think that they might have more opportunities despite their low education level. They overestimate the possibilities of promotion in society.

When this asymmetry is introduced, the reference income gets a coefficient which becomes strictly equal to that of mean individual income (and $\beta_2' \approx 0$). So there is now a perfect symmetry between the reference income and the individual permanent income, once we introduce an asymmetry in the perception of inequality. To summarise, income enters the life satisfaction equation by its short term transitory variation which has a positive influence (even if it is rather low) and by the ratio between long term income and reference income. If both are increased by the fraction, the effect is strictly neutral. We have managed to solve our empirical puzzle.

2.5.2 Identifying risk versus opportunity

The difference in attitude to inequality between the UK and Germany is still puzzling. We would like to investigate the attitude to inequality when it concerns others, which means inequality measured outside the educational reference group. We could try to measure inequality between educational

Table 2.6 – Estimation of a life satisfaction equation with Gini index

	Estimate	t value
Intercept	21.549	9.021
date2004	-0.034	-1.264
date2005	-0.117	-4.374
date2006	-0.044	-1.591
date2007	-0.033	-1.132
date2008	-0.011	-0.351
log(age)	-8.817	-7.048
log(age)^2	1.189	6.872
Min age	40.8	
marriage	0.482	13.263
log(adults)	-0.252	-6.983
log(1+kids)	-0.079	-2.592
health	-0.393	-30.126
$\Delta\log(y)$	0.050	2.103
$\log(\bar{y})$	-0.263	3.067
$\log(\bar{y}/y^r)$	0.394	-3.556
$Gini^r$	1.988	2.343
μ_1	0.585	15.563
μ_2	1.264	30.069
μ_3	1.990	45.774
μ_4	3.049	68.261
μ_5	4.461	94.483
σ	1.103	54.049
Log-likelihood	-24994.84	

groups, but this does not seem easy to implement. The other solution consists in measuring inequality within groups defined on another basis, such as regions. The BHPS provides a classification between 19 different regions: *Inner London, Outer London, South East, South West, East Anglia*, ... We can thus compute for each wave a Gini coefficient for each region which includes various education levels. We are looking for another measure of inequality which is independent of the human capital of the individual and thus this measure cannot be a measure of opportunity. The individual looks at the income distribution in his town, his neighbourhood. He looks at other people, not because they have the same education, but because they live broadly in the same place.

Of course, due to industrial specialisation there cannot be a clear inde-

Table 2.7 – Estimation of a life satisfaction equation with a Gini index for different educational groups

	Estimate	t value
Intercept	19.298	8.412
date2004	-0.036	-1.341
date2005	-0.124	-4.680
date2006	-0.054	-2.037
date2007	-0.048	-1.747
date2008	-0.028	-1.023
log(age)	-8.723	-6.919
log(age)^2	1.173	6.722
marriage	0.480	13.235
log(adults)	-0.250	-6.963
log(1+kids)	-0.077	-2.518
health	-0.394	-30.319
$\Delta\log(y)$	0.049	2.046
$\log(\bar{y})$	0.001	0.009
$\log(\bar{y}/y^r)$	0.129	3.060
Gini	1.750	2.061
Gini(lower)	0.658	3.859
μ_1	0.584	15.566
μ_2	1.263	30.084
μ_3	1.989	45.809
μ_4	3.049	68.334
μ_5	4.464	94.582
σ	1.103	54.204
Log-likelihood	-24976.92	

pendence between regions and education levels. However, when we reduce the education levels to 2 categories, the low educated versus the others, we find independence as a χ^2 test in a contingency table has value 27.54 with 18 DF and a P-value of 0.07. Aversion to inequality can be identified only if we restrict ourselves to the low educated group. This is what we find in Table 2.8. The regional Gini has a negative sign for the lower educated group, meaning that inequality within the region is perceived as a risk, but the effect is only significant at the 10% level. As a conclusion, lower educated people are both averse to global inequality on one side and on the other side over-estimate the possibilities they have within their educational group in term of future opportunities.

Table 2.8 – Estimation of a life satisfaction equation with Gini indices measuring risk and opportunity

	Estimate	t value
Intercept	19.980	8.851
date2004	-0.017	-0.658
date2005	-0.120	-4.550
date2006	-0.051	-1.894
date2007	-0.037	-1.365
date2008	-0.019	-0.696
log(age)	-8.855	-7.058
log(age)^2	1.191	6.854
marriage	0.477	13.181
log(adults)	-0.249	-6.925
log(1+kids)	-0.077	-2.525
health	-0.394	-30.291
$\Delta \log(y)$	0.048	2.041
$\log(\bar{y})$	0.001	0.010
$\log(\bar{y}/y^r)$	0.130	3.068
Gini-educ*(lower educ)	2.360	2.628
Gini-region*(lower educ)	-1.652	-1.873
μ_1	0.584	15.569
μ_2	1.263	30.097
μ_3	1.989	45.835
μ_4	3.049	68.375
μ_5	4.464	94.634
σ	1.103	54.307
Log-likelihood	-24976.89	

2.6 Conclusion

In this paper we have studied the relation between individual's income and individual's subjective well-being. In particular, we wanted to shed some light on the Easterlin paradox. Having access to panel data sets opens great possibilities, first to take into account individual effects and second to be able to introduce income dynamics. We could verify that the usual theory of adaptation is not sufficient (individuals get used to their income level and react only to variations of it, see Clark et al. 2008). Introducing long term income as an anchoring effect completed by short term variations provide an explanation for the level of well-being, but these variables become really significant only when a reference income is introduced.

A reference group is easy to define empirically. Considering only one sorting variable such as the education level is sufficient and additional variables do not fundamentally change the results. However, once the reference group is defined (we based it on a human capital definition), introducing the reference income is a much more complicated story as it leads to empirical puzzles. In particular, if we characterise the reference income only by its mean (or median), it appears that a rise in the reference income has to be compensated by a much higher rise in permanent income, by the order of several hundred percents. Or in other words if the position does not change, well-being decreases with long term income. This puzzle exists in the UK data, but not in the German data.

We managed to solve this empirical puzzle by considering a second characterisation of the reference income which is its dispersion, the income distribution inside each reference group, the income inequality inside the reference group. However, we had to consider an asymmetry of inequality perception between the low educated individuals and the others in order to solve the puzzle. We can conclude that the reference income is a key explanation for the Easterlin paradox, but that, at least for the UK data, the relation between the reference income and the level of well-being is very complex and highly non-linear.

Reference groups are not unique and can vary depending on the comparison purpose. In the same model, we can introduce several reference groups, provided they are independent, which means that they do not tell the same story. We could identify an aversion to overall inequality provided we restricted our attention to a particular group of individuals. It would have been interesting to justify more deeply our identification device, introducing for instance other attitude variables characterisation income expectations or the overall attitude to risk. This is left for a future research.

Chapter 3

A Bayesian subjective poverty line, one dollar a day revisited

3.1 Introduction

In different countries and at different times, the definition of poverty changes according to individual living situations and varying poverty perception. Even within a given society and at a given point of time, the critical level of income at which individuals are recognized as being poor is not perceived in the same ways by different income groups. The meaning of poverty differs between those groups as poverty can be, at least partly, a social construction. An example is given by preference drift (van Praag 1971, Goedhart et al. 1977).

Between different countries, the minimum basket of goods that ensures that physical and even mental well-being is not the same, just because living standards, traditions, habits and other social characteristics are different. The common view is that in the less developed countries, poverty is anchored to basic human needs, such as enough food, clean water, sanitation, clothing, shelter, health care and basic education. A poverty line in those countries is usually defined as an "absolute poverty line" that focus only on how much humans need for living, independently of the national income distribution. For richer countries, once the basic needs are satisfied, individuals tend to desire a more expensive basket of goods, e.g. more varied diets, suitable clothes, comfortable shelter, better health and higher education, just to be like the others and to be able to take a decent part in social life. The definition

0. This paper was co-authored with Michel Lubrano. It was presented at the 7th Rimini Bayesian Econometric workshop in May 2013 and at the 12th LAGV conference in Aix-en-Provence, June 2013. It appeared as an AMSE discussion paper No 2013-02.

of "poverty" in this case becomes more complex and is influenced largely by the perception of "economic inequality". An individual who considers himself as being poor may not face a problem of survival, but he is suffering from an envy comparison for what others have in his surroundings. The latter definition of poverty line is called a "relative poverty line".

Where could we put the limit between these two definitions of a poverty line? What is the list of countries which are considered as being sufficiently rich in order to afford a relative poverty line and what is the list of the other countries? Ravallion et al. (1991) showed that official national poverty lines varied little in comparison with mean consumption per capita for less developed countries, while above a critical level of mean consumption per capita, national official poverty lines had a much stronger elasticity with respect to consumption. Based on that previous finding, Ravallion and Chen (2001) and Ravallion and Chen (2004) proposed an international poverty line (a worldwide absolute poverty line) as being "\$ 1 per day" (\$1.08 at 1993 PPP). In a more recent paper, Ravallion et al. (2009) clearly identify two groups of countries in a new data set covering 74 developing countries with data collected within the period 1988-2005. They estimate a non-linear regression relating national official poverty lines to national mean consumption per capita, imposing a zero consumption coefficient for the group of less developed countries. With this model, the revised international poverty line (IPL) has risen to \$1.25 per day at 2005 PPP. Greb et al. (2011) re-revisited this study, using different econometric techniques and a different specification and found a somewhat higher international poverty line at \$ 1.45 per day. The difference between \$1.25 and \$1.45 is important. This adjustment means that 317.6 million extra people would fall in poverty in 2005. This gives us enough reason to revisit the problem of setting an international poverty carefully.

We are going to use Ravallion et al. (2009)'s new data set (as given at the end of their paper) and adopt a Bayesian approach in order first to take fully into account the model uncertainty and second to provide a posterior density for the obtained poverty line. We re-estimated the same empirical model used in Greb et al. (2011) and illustrate graphically where the difference lies between the two specifications. We then show how to define a poverty line as a function of the mean level of consumption of a reference group of less developed countries, the composition of that group being endogenously determined. We found in this latter analysis, an international poverty line eventually higher at \$1.48 per day. The posterior density of our international poverty line is not only much more concentrated but also within those intervals obtained using any of the previously mentioned specifications with a better fitness.

The paper is organized as follows. In section 2, we show how a subjective approach to poverty perception can be introduced, using macro cross-country data. With section 3, we develop the econometric techniques involved by a Bayesian approach to our problem. With section 4, we show how a Bayesian approach can illustrate some misspecification problems and provide a rational route to derive the posterior density of a world poverty line. Section 5 concludes.

3.2 Poverty lines and preference drift

What is the elasticity between a poverty line with respect to average income/consumption level? That elasticity is equal to zero by definition in the case of absolute poverty lines while it is equal to one in the case of relative poverty lines. We obtain an elasticity in between with subjectively determined poverty lines. How is it possible to build an international subjective poverty line with a suitable income elasticity? This is the question we shall try to answer in this section.

3.2.1 Subjective poverty lines

A poverty line can be defined on an individual subjective basis. We refer to the minimum income question (MIQ) that can be found for instance in Kapteyn et al. (1988) and which is phrased as follows: *what is the minimum income that you would need in order to make the two ends meet?* If z_i is the reported answer, y_i the actual income of the household and x_i a vector of characteristics of the household (such as its composition), then the following regression [1] is estimated over all the individuals:

$$z_i = \alpha + \beta y_i + \gamma x_i + \epsilon_i. \tag{3.1}$$

An estimated subjective poverty line for a single country corresponds to a fixed point for every type of household composition x:

$$z^* = \frac{\hat{\alpha} + \hat{\gamma} x}{1 - \hat{\beta}}. \tag{3.2}$$

With a fixed point, we classify as being poor households having reported an income below the answer they gave to the MIQ and that for each type of household composition. [2]

1. A log-log regression is also possible as in Van den Bosch et al. (1993).
2. Other definitions and methods to establish a subjective poverty line were proposed in the Literature. For instance, the Leyden poverty line as introduced in Goedhart et al.

Van den Bosch et al. (1993) used this approach to compare poverty in European countries and compared their estimated poverty lines to the official ones determined as 50% of the mean national disposable income. In most cases, the subjective poverty line is over the official figure. A preference drift, at least in developed countries is clearly identified, see for instance van Praag (1971) for Belgium. That means, that individuals do not have the same perception of poverty, depending on their income. The perceived minimum necessary income rises with the level of personal income. This is the reason why some methods like the CSP (see footnote 2) consider only one part of the sample in order to compensate for preference drift.

Pradhan and Ravallion (2000) argued that the subjective approach based on the MIQ is not a valid practice for developing countries. The reason is quite simple to understand. The notion of income is subject to huge variations in developing countries because the consumption of some items might not require access to a market. Pradhan and Ravallion (2000) develop another approach which is based on a different set of questions. Empirical surveys led by the World Bank have introduced a question about the adequacy of household consumption (less than adequate, adequate, more than adequate) for five different items (food, housing, clothing, health, schooling). For each item, an ordered probit model can be adjusted to confront these answers to the level of actual consumption and various indicators about household composition. A poverty line is determined for each item and then these five items are aggregated to form an overall subjective poverty line. In Pradhan and Ravallion (2000), a particular variable appears to be particularly important for the Nepal equation (not for Jamaica) which is the log mean consumption of the neighbouring area. The is evidently a comparison effect. The vision that households have of poverty not only depends on the consumption of a basket of goods, but also on what the other households consume.

3.2.2 Absolute poverty lines for developing countries

We now consider the case of several countries. This means that individuals are replaced by countries and that the country becomes the unit of observation. Equation (3.1) becomes

$$z_j = \alpha + \beta C_j + \epsilon_j \tag{3.3}$$

(1977) which is based on the Income Evaluation Question and the estimation of a particular social welfare function. We can also mention the CSP (Center for Social Policy) of the University of Antwerp introduced in Deleeck (1989) and reviewed in Van den Bosch et al. (1993) which relies both the MIQ and a supplementary question concerning financial ease. Only the households answering *with some difficulty* to that second question are taken into account to define the poverty line.

where z_j is the national official poverty line and C_j is the mean consumption level per capita in country j. The same fixed point algorithm provides the level of the common poverty line:

$$z^* = \frac{\alpha}{1-\beta}.$$

Ravallion et al. (2009) assumed that for very poor countries, the restriction $\beta = 0$ has to be imposed. With that restriction, the common poverty line z^* becomes the empirical mean of the z_j and its elasticity with respect to national consumption is zero.

Working on a data set of 74 developing and developed countries, Ravallion et al. (2009) identified a group of 15 very poor countries with an average level of private consumption per capita of less that \$60 a month. They estimate model (3.3), introducing an exogenous break at $C_j = \$60$:

$$z_j = \mathbb{I}(C_j < \$60)\alpha_1 + \mathbb{I}(C_j > \$60)(\alpha_2 + \gamma C_j) + \epsilon_j \tag{3.4}$$

where z_j is an official poverty line in PPP dollars, C_j the average level of private consumption per capita in PPP dollars and $\mathbb{I}(.)$ an indicator function equal to one when the condition in parenthesis is satisfied and zero otherwise. The estimated poverty line is given by the regression coefficient α_1. This coefficient represents an estimate of the empirical mean of the z_j computed for the countries for which $C_j < \$60$. This group of countries is said to represent the reference group to compute the poverty line. The sole common information between the two regimes in (3.4) is that they share a common variance for the residual term.

3.2.3 Evidence of preference drift among developing countries

The main assumption made in Ravallion et al. (2009) and all the related work is that of an absolute poverty line for the poorest countries. When looking at the figures reported in the data base of Ravallion et al. (2009), we however do find a relation between z_j and C_j for the group of very poor countries, even if that relation is not of the same amplitude as for richer countries. For countries with a mean C lower than \$60 a month, the poverty line represents on average 92% of the mean consumption level while that factor drops down to 45% for the richer group of countries. This last figure is much more in accordance with the usual definition of a relative poverty line which corresponds usually to half of the mean income. The first figure of 92% can be completed by computing the average of the reported poverty lines . We find \$38 with a standard deviation of \$12. How could we explain such a large standard deviation? In this group the minimum and maximum poverty lines are \$19 and \$59.

Rather than being fixed, a poverty line for a less developed country has a certain elasticity with respect to income or consumption. There is a relation, even if it is of a different nature, between the average level of consumption and a reasonable poverty line, which is not necessary the official one. To illustrate that point, we would like to report the controversy that took place recently in India around the decision of the Indian Government to reduce the level of the official poverty line, following the recommendation of the Tendulkar commission Tendulkar (2009). It is commonly admitted that the necessary amount of calories is around 2100 (Ravallion 1994). The Indian Government decided that 1770 calories were enough because on average individuals did consume that level in India. The official rate of poverty was thus reduced to Rs 28.65 per capita for daily consumption in cities and to Rs 22.42 in rural areas. The objective was to reduce the official poverty rate which went down to 29.8% with these new figures (note that the World Bank estimated the rate of poverty in India to be 32.7% in 2010 with a poverty line of \$1.25 a day). The Indian press reported large protestation, which can be understood when we know that India is a very fast developing country (comparison effect). But certainly inequality is also rapidly increasing. This decision to reduce the official poverty line is at odds with what was claimed in other parts of the commission report, that "*Fundamentally, the concept of poverty is associated with socially perceived deprivation with respect to basic human needs*". So there is a large gap between what society perceives and what the official agencies publish. Another example can be found in British history. The poverty line used in the nineteen century was defined with respect to a consumption basket meant to provide the necessary calories to survive. Atkinson (1983, p. 188, chap. 10) recalls the example of English workers who went on strike because tea was planned to be withdrawn from the official basket of goods for computing the poverty line. Despite the fact that tea had no nutritional value, it had a social value. The composition of that basket was socially determined.

This discrepancy between an official poverty line and what individuals perceive is of course not specific to developing countries. In Van den Bosch et al. (1993), we see that in countries of Southern Europe, a subjective poverty line is also much higher than the official line. Our model should take into account the relations existing between z_j and C_j in order to define a world poverty line.

3.2.4 An international subjective poverty line

The poverty line which is proposed both in Ravallion et al. (2009) and Greb et al. (2011) consists in computing the mean poverty line of a reference group when that reference group is given exogenously in Ravallion et al. (2009) or is endogenously determined in Greb et al. (2011). The idea we would like to illustrate here is that a poverty line for the less developed countries has still to be determined as a function of the characteristics of a reference group, but that this poverty line should depend also on a reference income (or a reference consumption level). Consequently, the model that we shall estimate is

$$z_j = s_j(\alpha_1 + \gamma_1 C_j) + (1 - s_j)(\alpha_2 + \gamma_2 C_j) + \epsilon_j \quad (3.5)$$

$$s_j = \begin{cases} 1 & \text{if } C_j < \theta \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

where θ is an unknown threshold. The new poverty line will be determined as a conditional expectation

$$\mathrm{E}(z_j | s_j = 1) = \alpha_1 + \gamma_1 \mathrm{E}(C_j | s_j = 1). \quad (3.7)$$

In words, the poverty line we propose for less developed countries is a function of a reference group consumption level which is taken to be equal to the mean consumption of that reference group. It is different from a usual relative poverty line in the sense that it depends not on the national mean consumption per capita but on the mean consumption of a more general group, called the reference group. We call this new poverty line a subjective poverty line not because it depends on subjective data, but because it relates to a common group to which countries are supposed to identify themselves. They judge their poverty line by reference to that group. The notion of a reference group appeared in the happiness economic literature as a possible explanation to the Easterlin paradox. Individual satisfaction is a function not mainly of the level of income, but of the difference between their income and a reference income which is taken as the mean income of the reference group. See for instance Ferrer-i-Carbonell (2005) for an empirical investigation. We try here to translate that concept to countries and poverty lines.

The convenient way both to determine an estimate of the threshold parameter θ and to take into account the uncertainty in the determination of the reference group is to adopt a Bayesian approach, as we shall see in the next section.

3.3 Bayesian inference for regression models with a break

The generic model we want to estimate is a two regime model with a break determined when a variable w_i is lower or greater than an unknown threshold θ:

$$\begin{aligned} \mathrm{E}(y_i|x_i) = & \; x_i'\beta_1 \quad \text{if } w_i < \theta \\ \mathrm{E}(y_i|x_i) = & \; x_i'\beta_2 \quad \text{if } w_i > \theta. \end{aligned}$$

y_i is the dependent variable, x_i a set of exogenous variables and w_i is the regime shift variable which is supposed to be exogenous or predetermined. θ is a threshold parameter. We introduce the unobserved variable s_i defined as:

$$s_i = \begin{cases} 1 & \text{if } w_i < \theta \\ 0 & \text{otherwise.} \end{cases}$$

The regression model that we shall consider is

$$y_i = s_i x_i'\beta_1 + (1 - s_i)x_i'\beta_2 + \epsilon_i,$$

where the error term ϵ_i is supposed to be normal with zero mean and constant variance σ^2. For inference purposes, it is useful to define the following matrix:

$$X(\theta) = [s_i x_i', (1 - s_i)x_i']. \tag{3.8}$$

Thus we have the more compact form:

$$y = X(\theta)\beta + \epsilon, \tag{3.9}$$

where y is a vector containing the N observations of y_i and β the vector containing parameters β_1 and β_2.

3.3.1 Likelihood and posteriors

Considering N observations, the likelihood function of model (3.9) is:

$$L(\beta, \sigma^2, \theta; y) \propto \sigma^{-N} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[y_i - X_i'(\theta)\beta]^2\right]. \tag{3.10}$$

Conditionally on θ, this is the likelihood function of a usual regression model, so that natural conjugate prior densities for β and σ^2 belong the normal inverted gamma2 family:

$$\begin{aligned} \pi(\beta|\sigma^2) &= f_N(\beta_0, \sigma^2 M_0^{-1}), \\ \pi(\sigma^2) &= f_{Ig}(\sigma^2|\nu_0, s_0). \end{aligned}$$

The conditional posterior densities of β and σ^2 are:

$$\begin{aligned} \pi(\beta|\theta, y) &= f_t(\beta|\beta_*(\theta), s_*(\theta), M_*(\theta), \nu_*), \qquad (3.11) \\ \pi(\sigma^2|\theta, y) &= f_{Ig}(\sigma^2|\nu_*, s_*(\theta)), \qquad (3.12) \end{aligned}$$

where

$$\begin{aligned} M_*(\theta) &= M_0 + X'(\theta)X(\theta), \\ \beta_*(\theta) &= M_*^{-1}(\theta)[X'(\theta)y + M_0\beta_0], \\ s_*(\theta) &= s_0 + \beta_o^{'}M_0\beta_0 + y'y - \beta_*'(\theta)M_*(\theta)\beta_*(\theta), \\ \nu_* &= \nu_0 + N. \end{aligned}$$

The posterior density of θ is proportional to the inverse of the integration constant of the Student posterior density (3.11) times the prior density of θ:

$$\pi(\theta|y) \propto |s_*(\theta)|^{-(N-k)/2}|M_*(\theta)|^{-1/2}\pi(\theta). \qquad (3.13)$$

As there is no conjugate prior for θ, we are free to use any form of parametric density. A convenient choice is to use a uniform prior between bounds or a non-informative prior. The marginal posterior densities of β and σ^2 have to be found using numerical integration as we have:

$$\pi(\beta|y) = \int f_t(\beta|\beta_*(\theta), s_*(\theta), M_*(\theta), \nu_*)\pi(\theta|y)d\theta, \qquad (3.14)$$

and

$$\pi(\sigma^2|y) = \int f_{Ig}(\sigma^2|\nu_*, s_*(\theta))\pi(\theta|y)d\theta. \qquad (3.15)$$

The dimension of θ being one, we could use a traditional deterministic integration rule, like the Simpson rule in order to evaluate these densities. However, as we will be interested in transformations of the parameters, a simulation method is better. [3] As (3.13) is a marginal density, we have simply to find a feasible grid over which to evaluate it, compute numerically the cumulative and then use the inverse transformation method to draw a value for θ. Briefly, the grid over which to evaluate (3.13) has to be chosen carefully. It should cover most of the probability, but it should also avoid identification problems as detailed in Bauwens et al. (1999, p. 235). The grid should be chosen in such a way that there are enough observations per regime. The

3. It is very easy to compute the posterior density of a transformation of a parameter when we have posterior draws from this parameter. We just have to take the transformation of each draw as draws from the posterior of the transformed parameter. Using deterministic integration rules leads to much more complicated solutions.

domain of definition of θ is given by [Min(w_i), Max(w_i)]. But its bounds cannot be reached, because otherwise, the model would not be identified in the absence of prior information. We draw a value of θ from $\pi(\theta|y_i)$. Using this draw, we draw a β from the conditional posterior $\pi(\beta|\theta, y)$ given in (3.14) which is a Student density.

3.3.2 The two variance case

For modeling purposes, it will be useful to consider the possibility of having different variances in the two regimes. The consideration of heteroscedasticity is also economically interesting that in this study we could expect that for richer countries the determinants of poverty line should be more complex or culture dependent. This would lead to a greater noise.

We keep the same dichotomous variable s_i as in the original model and assume this time that:

$$\text{Var}(\epsilon_i) = s_i\sigma_1^2 + (1 - s_i)\sigma_2^2 = \sigma_2^2(s_i\phi + 1 - s_i) = \sigma^2 h_i(\theta, \phi), \tag{3.16}$$

as detailed in Bauwens et al. (1999, p. 236). Let us now divide the observations by $\sqrt{h_i(\theta, \phi)}$ in order to get a regression model with homoskedastic errors of variance σ^2:

$$\begin{aligned} y(\theta, \phi) &= [y_i/\sqrt{h(\theta, \phi)}], &(3.17)\\ X(\theta, \phi) &= [s_i x_i'/\sqrt{h(\theta, \phi)}, (1 - s_i)x_{it}'/\sqrt{h(\theta, \phi)}]. &(3.18)\end{aligned}$$

The regression model becomes:

$$y(\theta, \phi) = X(\theta, \phi)\beta + \epsilon,$$

its likelihood function being

$$\begin{aligned} L(\beta, \sigma^2, \theta, \phi; y) \propto\ & \sigma^{-N} \prod_{i=1}^{N} h_i(\theta, \phi)^{-1/2} \times \qquad (3.19)\\ & \exp\left[-\frac{1}{2\sigma^2}(y(\theta, \phi) - X(\theta, \phi)\beta)'(y(\theta, \phi) - X(\theta, \phi)\beta)\right].\end{aligned}$$

The conditional posterior densities of β and σ^2 are the same as before. We just have to replace y and $X(\theta)$ by $y(\theta, \phi)$ and $X(\theta, \phi)$ in the necessary expressions. The joint posterior density of θ and ϕ has the form:

$$\pi(\theta, \phi|y) \propto \prod_{i=1}^{N} h_i(\theta, \phi)^{-1/2}|s_*(\theta, \phi)|^{-(N-k)/2}|M_*(\theta, \phi)|^{-1/2}\pi(\theta)\pi(\phi). \tag{3.20}$$

It is slightly more difficult to draw θ and ϕ jointly from this bivariate density (3.20) than to draw θ from the univariate density (3.13). It is always possible in theory to decompose a bivariate density into

$$\pi(\theta, \phi|y) = \pi(\phi|\theta, y) \times \pi(\theta|y),$$

so that we first draw in the marginal density $\pi(\theta|y)$ and then in the conditional $\pi(\phi|\theta, y)$. To apply this method, we have first to determine a grid over θ and ϕ in order to fill up a matrix. From this matrix of points, we can determine numerically the marginal density $\pi(\theta|y)$. For a given draw of θ, we have to find the corresponding conditional $\pi(\phi|\theta, y)$. Of course, we will not have a draw of θ that corresponds exactly to a line of the initial matrix of points. So we shall have to proceed by linear interpolation between two lines as explained in the Appendix A.2.1.

3.4 Data and estimation

The data come from Ravallion et al. (2009) who have considered 74 developing countries. The data set includes national official poverty lines (PL) (or academic poverty lines in some cases) and Private Consumption Expenditures (PCE) per capita. These data report to different years from 1988 to 2005. They have been adjusted by the household consumption PPP's collected during the international comparison program of 2005 (World bank, 2008). The PCE and PL variables are reported on a monthly basis. This data set is an improvement over the old data set used in Ravallion et al. (1991) which covered only 33 countries and had a weaker price adjustment. [4]

3.4.1 Revisiting the initial model

Ravallion et al. (2009) estimate (3.4) while Greb et al. (2011), using the same data set, estimate a slightly different model

$$z_j = s_j\alpha_1 + (1 - s_j)(\alpha_2 + \gamma \log C_j) + \epsilon_j, \tag{3.21}$$

where $s_j = 1\!\!\mathrm{I}(C_j < \theta)$. The reference group is endogenously determined by θ which is now estimated, instead of chosen fixed on a priori grounds as in Ravallion et al. (2009).

4. We must note however that this data set is not exempt of oddities. For instance, the official poverty line for intermediate urban areas in Senegal was 661.7 CFA in 2005, which made \$1.06 at the current rate of exchange while when using the PPP, the official poverty line drops to \$0.64.

Both models (3.4) and (3.21) include only a constant term in the first regime. They differ mainly because Ravallion et al. (2009) adopt a formulation in levels while Greb et al. (2011) prefer to use a formulation in logs. Using a Bayesian approach provides us the adequate tools to discuss and compare those two alternative specifications. When we estimate both formulations with an unknown threshold θ, we observe that the model in levels provides a rather inprecise estimation for θ as we have $\mathrm{E}(\theta|y) = 98.77$ and a large standard deviation of (44.36) while the model in logs provides a much higher value for θ as $\mathrm{E}(\theta|y) = 138.99$, but also a much smaller standard deviation of (34.78). As a consequence, the poverty line is better estimated with the model in logs. This is again well apparent if we examine the posterior density of θ in both models as displayed in Figure 3.1. We see that in order

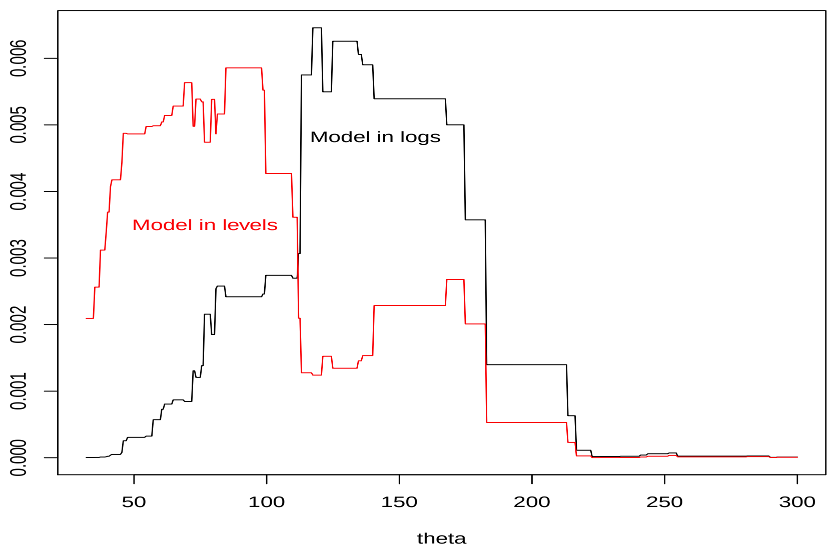


Figure 3.1 – Why a model in logs is better

to deliver a reasonable message the model in levels has to be equipped with a strong prior on θ in order to limit its range to the first mode of its posterior density, say a range of [32,120]. The posterior corresponding to the model is logs is uni-modal, delivering thus a single message and needs a much less informative prior for θ.

We report in Table 3.1 estimation results for model (3.21) using a non-informative prior and 5 000 draws. An estimate for a World Poverty Line is obtained by re-scaling the posterior density of α_1, considering $\alpha_1/365 * 12$. The corresponding graph is given in Figure 3.2. A 90% confidence interval

Table 3.1 – Bayesian inference for initial model

	Estimate	std.error
α_1	43.44	7.02
α_2	-467.95	109.44
γ_2	104.45	10.56
θ	138.99	34.78
σ^2	1425.70	246.09
n	74	
BIC	754.07	
Estimated IPL	1.43	0.23

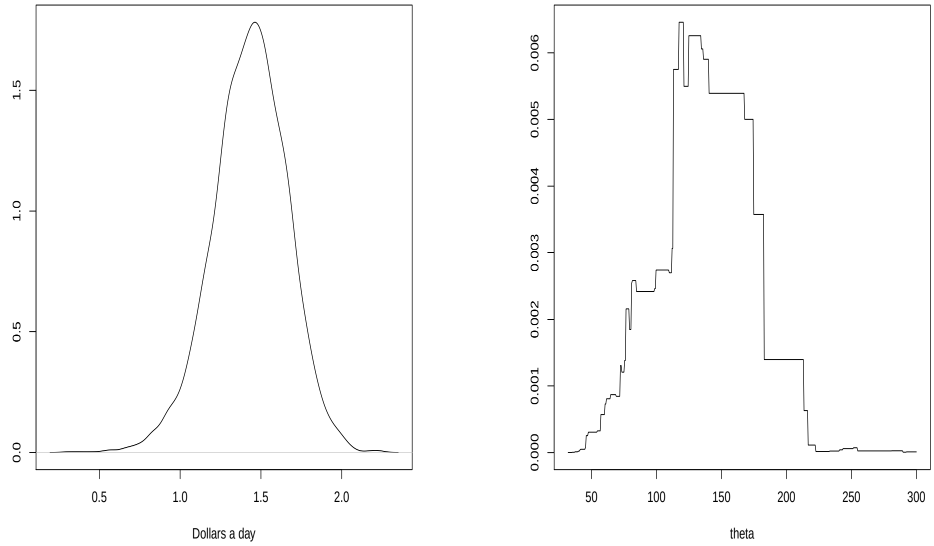
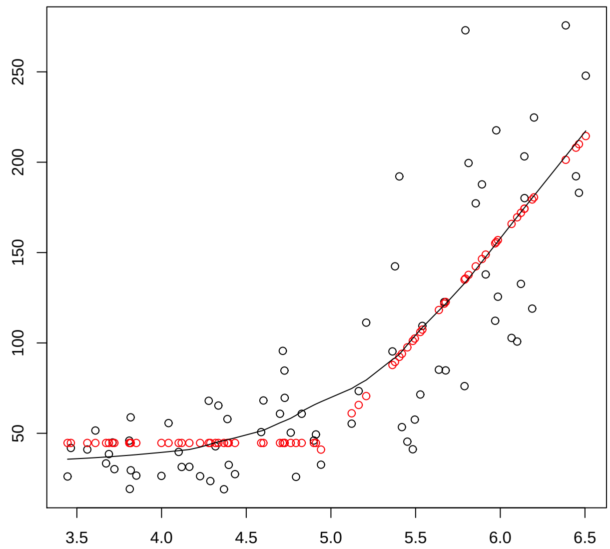


Figure 3.2 – Rescaled posterior density of α_1 and posterior density of θ

is [1.05, 1.79]. This result is similar to that reported in Greb et al. (2011) who found a 90% confidence interval of $[1.10, 1.72]$. Of course the Bayesian posterior interval is slightly larger (also underlined in Hansen (2000), the distribution of the estimated threshold θ is not standard (see Figures 3.1 and 3.2) and thus using an asymptotic approximation as in Greb et al. (2011) is not the right method to report the empirical uncertainty. A Bayesian approach provides the small sample distribution of θ and thus allows to take into account uncertainty in the determination of an empirical poverty line in a rational way. Whatever the estimation method, (3.21) leads to the determination of a much larger reference group than that obtained in Ravallion et al. (2009). We have on average 38 countries when there were only 15 in Ravallion et al. (2009) where θ is fixed and equal to \$60. This increase in the size of the reference group leads to a slightly larger value for the poverty line.

A graph of the predictive density, as reported in Figure 3.3, suggests that the variance of the error term is not the same in the two regimes. A



The red circles represent the predictive curve, the black circles are the data points. The solid line represent the non-parametric estimate of regression (3.3).

Figure 3.3 – Posterior fit of the initial model in logs

model with two variances is even more coherent with the theoretical model of Ravallion et al. (2009) as in the first regime, the poverty line is supposed to be constant and the level of per capita consumption rather low. We thus consider the alternative model:

$$z_j = s_j\alpha_1 + (1 - s_j)(\alpha_2 + \gamma_2 \log(C_j)) + s_j\epsilon_{1j} + (1 - s_j)\epsilon_{2j}. \tag{3.22}$$

Table 3.2 validates the existence of two different error term variances as their ratio ϕ is much lower than 1. α_1 is estimated in a much more precise way with a standard deviation that goes down from 7.02 to 3.93, leading to a narrower 90% confidence interval of [1.20, 1.62] for the poverty line also with a better fitness of the model. This indicates that the consideration of unequal variances in the lower and upper part of the data set is essential to

Table 3.2 – Bayesian inference the two variance case

	Estimate	std.error
α_1	43.16	3.93
α_2	-460.00	135.92
γ_2	103.16	9.99
θ	144.48	32.37
σ_1^2	312.06	100.4
σ_2^2	2518.2	647.4
ϕ	0.13	0.050
n	74	
BIC(evaluated at 1st mode of θ)*	725.39	
BIC(evaluated at mean θ)	721.75	
BIC(evaluated at 2nd mode of θ)	719.98	
Estimated IPL	1.42	0.13

The first modal locates at $\theta = 100$ and the second modal locates at $\theta = 160$.

our analysis. The lack of such consideration would certainly bias the estimate of parameter θ thus biases also the estimate of IPL.

However, the posterior density of θ becomes bimodal (Figure not given here) which leads us to look for a better model.

3.4.2 Preference drift

In the approach of both Ravallion et al. (2009) and Greb et al. (2011), the assumption is that for low income countries, the poverty line should have an absolute definition, which means that it is independent of income or consumption. When we look at Figure 3.3, we see that this assumption is not fully coherent with the data. In the first regime, the official poverty line seems to depend on the level of consumption, however with a much lower slope than in the second regime. We shall now estimate our preferred model (3.5) but including two variances so as to obtain:

$$z_j = s_j[\alpha_1 + \gamma_1 \log(C_j)] + (1 - s_j)\,[\alpha_2 + \gamma_2 \log(C_j)] + s_j\epsilon_{1j} + (1 - s_j)\epsilon_{2j}.$$

In Table 3.3, we have reported two versions of this model. Apparently we must have a parsimonious parameterization in the first regime and suppress the constant term if we include log consumption. The final model is that corresponding to the second panel of this table. Compared to the initial model in logs as reported in Table 3.1 in term of BIC, we have a much better fit. For the first regime which is of direct interest to us, the variance of the error

Table 3.3 – Model with preference drift and two variances

	Estimate	std.error	Estimate	std.error
α_1	-26.62	26.87	-	-
γ_1	16.92	2.40	10.63	0.67
α_2	-497.8	151.9	-496.2	142.9
γ_2	110.2	3.85	109.1	3.81
θ	172.0	19.29	167.2	18.24
σ_1^2	284.6	69.93	297.92	74.59
σ_2^2	2780.4	737.0	2668.0	692.8
ϕ	0.11	0.038	0.12	0.041
BIC	716.36		716.16	

term is much lower. The posterior density of θ is much more concentrated, indicating a much more precisely determined sample separation. The posterior density of γ_1, which is of direct interest to us now, is much more concentrated than the posterior density of α_1 (which was of prime concern) in the initial model. In our model, $\log(C_j)$ does have an influence in determining the official national poverty line, but its impact is ten times lower than what it is in the second regime ($\gamma_1 \ll \gamma_2$).

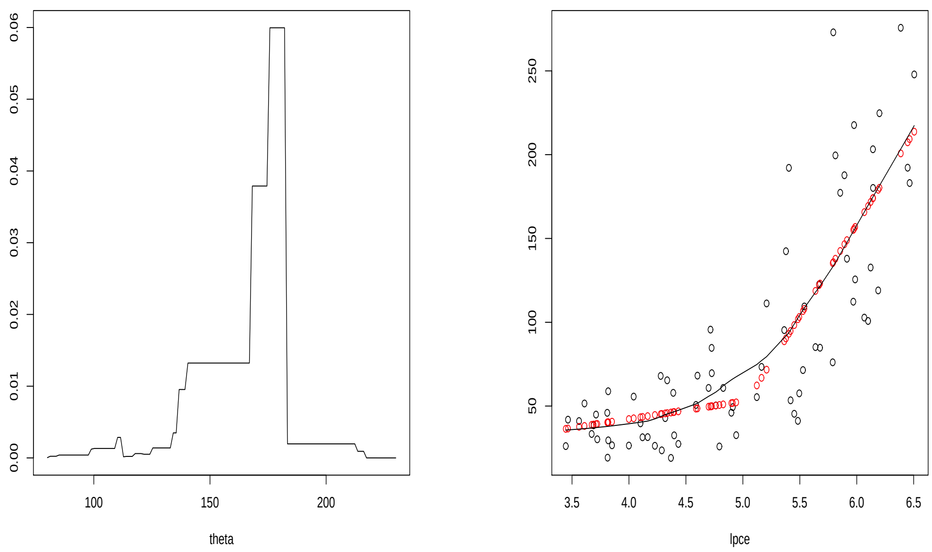


Figure 3.4 – Posterior density of θ and posterior fit of the last model

We have computed the posterior mean probability that a country belongs to the reference group. We have 26 countries for which this probability is equal to 1: Bangladesh, Benin, Burkina-Faso, Cambodia, Chad, Congo-Rep, Ethiopia, Gambia, Ghana, Guinea-Bissau, Malawi, Mali, Mongolia, Mozambique, Nepal, Niger, Nigeria, Rwanda, Senegal, Sierra-Leone, Tajikistan, Tanzania, Uganda, Vietnam, Yemen, Zambia. In this first group the maximum consumption is \$81 for Vietnam (to be compared to the \$60 of Ravallion et al., 2009). With a 95% probability bound, we get a group of 38 countries with the following 12 additional members: Cameroon, China, Cote d'Ivoire, Djibouti, India, Kenya, Kyrgyz, Lesotho, Mauritania, Moldova, Pakistan and the Philippines.

3.4.3 How to simulate the posterior density of the IPL

We have to find a posterior density for the IPL, based on the first regime characteristics. It is obtained as a transformation of the parameters. Following (3.7), we define the IPL as being in our case:

$$IPL = \gamma_1 \mathrm{E}(\log(C_j)|C_j < \theta).$$

In order to simulate the IPL, we must have draws of γ_1 and θ. For each draw of θ, we determine the corresponding reference group and compute a value for the sample mean of $\log(C_j)$. We then multiply this value by the corresponding draw of γ_1. The algorithm is as follows. We have stored draws from γ_1 and θ noted γ_1^i and θ^i.

1. Start a loop in i
2. Given θ^i, determine a sample separation and n_1^i the sample size in the first regime
3. Compute $z^i = \gamma_1^i \sum_{k=1}^{n_1^i} \log(C_k) 1\!\!\mathrm{I}(C_k < \theta^i)/n_1^i$
4. store z^i
5. End loop

We get a mean IPL of 1.48 dollars a day with a standard deviation of 0.096. A 90% confidence interval is [1.32-1.64]. We have given in the previous subsection a list of 26 countries which had a probability equal to 1 to be included in the group of reference. The posterior density of the IPL when the reference group is limited to this group is obtained as a simple transformation of the posterior density of γ_1. We get a mean IPL of 1.39 dollars a day (0.086). We give in Figure 3.5 a graph of the posterior density of these two possible IPL. We have also added the posterior density of the poverty line using the first model in logs as defined in Greb et al. (2011) as well as the poverty line corresponding to the approach of Ravallion et al. (2009). For this last option,

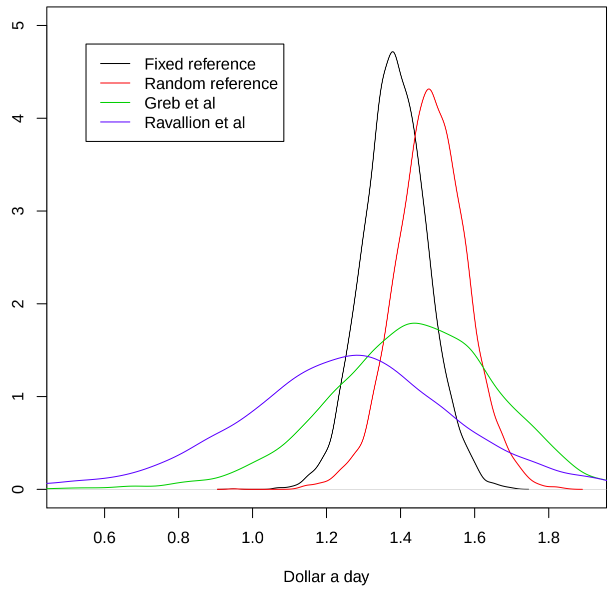


Figure 3.5 – Posterior density of four poverty lines

we had to specify a prior information θ which was compatible with range of θ representative of the approach of Ravallion et al. (2009), i.e. $\theta \in [32, 120]$, in order to eliminate the secondary mode in the posterior density of θ. We get a mean poverty line of \$1.26 (0.33), compared to the \$1.25 found in Ravallion et al. (2009).[5]

The four poverty lines presented in Figure 3.5 illustrate four different possible approaches.

1. The first poverty line of the plot corresponds to a fixed reference group. That group is used to compute a reference consumption level, common to that group of countries (however determined by the model). The common poverty line is defined as proportion of this reference consumption level. Uncertainty comes from that proportion γ_1. It corresponds to a sample-based prior for θ.

2. The second poverty line has a slightly higher posterior mean as it corresponds to a larger reference group, but it takes into account the whole uncertainty of the model. This is our proposed poverty line.

5. If we had restricted the prior range of θ to [32-60] so as to follow more closely the options of that paper, we would have obtained a poverty line of \$1.23 with a larger standard deviation (0.44) and a reference group of 13 countries instead of 22.

3. The last two poverty lines obey a different philosophy. They measure simply the mean of different national poverty lines, using each a particular reference group. In one case, we use a model formulated in logs, which was shown to correspond to a better specification. In the other case, we use a model specified in levels which corresponds to the initial model of Ravallion et al. (2009). Both of them have a much larger standard deviation.

With a poverty line of $1.25, following the data published in Chen and Ravallion (2008), we have 1.4 billion poor people in the developing World. With a mean poverty line of $1.48 as we found, this figure goes up to more than 1.7 billion and the headcount index passes from 25.7 to 31.5.[6]

3.5 Conclusion and comments

Defining an international poverty line is an important objective, because it allows to identify the countries where poverty is located and it leads to measure the number of poor people in the developing world. Knowing these numbers and the localization of poverty, it is easier to devise anti-poverty policies and to evaluate the results of those policies later on.

We have seen in this paper that it is not an easy task to devise a poverty level. The one dollar a day line had to be reformed and Ravallion et al. (2009) was a major attempt to do this. Their newly proposed poverty line is the lowest of the different poverty lines we have reviewed. They all rely on a different definition of a reference group.

We have shown that a large uncertainty is attached to the calculation of poverty lines (mainly the uncertainty of the distribution of θ). The model specification (level versus log and equal variance versus two variances) and the definition of poverty line (absolute versus relative) also lead to different results.

The final mean poverty line we obtain (1.48 dollar a day) is larger than the 1.25 of Ravallion et al. (2009). But it is well in a reasonable confidence interval along with a better goodness of fit. The posterior density of our poverty line covers all the other proposed point alternatives. Due to the way it is computed, our poverty line compels to the logic of an international subjective poverty line.

The final point we would like to make concerns the Bayesian approach. We have used quite standard tools, even if they could seem complex for a reader

6. In fact the figures given in Chen and Ravallion (2008) are for a range of poverty lines which are $1.00, $1.25,$1.45, $2.00 and $2.50.

not familiar with the field. With these tools we have visualized the origin of some questions concerning the model to be used and its specification. The posterior density of the break point θ was particularly useful in this respect. And finally, we could compare various assumptions concerning the determination of an international poverty line.

Chapter 4

Globalization, income mobility and the preference for redistribution

4.1 Introduction

Bourguignon (2012) has pointed out that globalization has reduced inequality between countries, but increased inequality within countries. More precisely, due to the weight of rising China and India, world inequality started to decrease in 2000 and that movement went on since that date. Meantime and starting from 1980 for most OECD countries including Sweden, within country inequality has constantly risen (see OECD 2011). Because globalization has winners and losers, many economists tried to study the consequences of economic openness on the welfare state and on the demand for redistribution. Building on earlier works, Rodrik (1998) developed the idea of a *compensation hypothesis*. Globalization is synonymous of a serious risk increase for the unskilled workers. The government compensates for that risk by increasing public spending. Walter (2010) tests this assumption on Swiss data and found an empirical validation for it.

What are the mechanisms behind this? Within each country, economic growth benefits the different segments of the population unequally, creating a dynamics for both inequality and new opportunities. We have the example of the large increase in wage inequality which is well documented in the U.S. literature (see e.g. Bound and Johnson 1992, Katz and Murphy 1992 or Murphy and Welch 1992). More precisely globalization creates a movement in income

---

0. This paper was co-authored with Michel Lubrano. It was presented at the 13th LAGV conference in Aix-en-Provence, June 2014 and IIPPE, Naples, Sep. 2014

dynamics and in income distribution with the following characteristics:

- Globalization is an opportunity for high skilled workers because developed countries have access to a wider market for high technology products.
- Globalization is an opportunity for high wage earners such as superstars and Chief Executive Officers. As shown in Gabaix and Landier (2008), globalization leads to higher top wages because in tournaments and in superstars models, the prize is higher when the market is larger.
- Globalization is a risk for low skilled workers because they are in competition with low skilled workers from less developed countries. With delocalizations, there is an increase in unemployment risk and a general pressure on low wages, a pressure which is amplified by fiscal competition.

The aim of the present paper is to explicit the relation between income dynamics and the demand for redistribution in a context of globalization. The classical model of Meltzer and Richard (1981) determines an optimal taxation rate by inspecting the gap between the median and the mean incomes. This is a static model based on the theory of the median voter of Romer (1975). Individuals who have an income below the mean will vote for redistribution because they expect to receive more than their tax contribution. This model predicts that an increase in inequality will induce an increase in demand for redistribution. Although it has received a recent renewal of interest with Karabarbounis (2011), this model is too simple, just because it is a static model that cannot capture income dynamics. If the same voters anticipate that their future income will be greater than the mean, their future status will be changed from being net tax receivers to being net tax payers. Consequently, they will vote against redistribution. Benabou and Ok (2001c) formalized this idea with the POUM hypothesis or Prospect of Upward Mobility. They clearly explicit a relation between income dynamics, demand for redistribution and risk aversion. However, the main result of the paper (some voters under the mean income can vote against redistribution) relies on specific assumptions concerning regularity of the income dynamics and population homogeneity. When slightly relaxing these two assumptions, the existence of a steady state vanishes. In a heterogeneous population, preference for redistribution can no longer be analysed by simply considering the properties of the income dynamics and of its steady state. The properties of the income dynamics have to be confronted to declared preferences. We have thus to depart from usual empirical studies like Alesina and La Ferrara (2005) which analyse cross country surveys such as the GSS (General Social Survey) or the ISSP (International Social Survey Program). We have to make use of an integrated panel survey where individual data are available both for income dynamics and opinions concerning redistribution (even if the later variables are not as detailed as those available in the above mentioned social surveys). We shall base our investigations on the BHPS (British Household Panel Survey).

The paper is organized as follows. In section two, we build a toy model of income redistribution using the lognormal distribution. This model is designed to show the role of each of the assumptions made in Benabou and Ok (2001c). We detail how to relax those assumptions and in particular the first order stochastic dominance assumption and its interaction with risk aversion. In section three, we describe our data set (BHPS) concerning job status, wages and preference for redistribution and precise our empirical strategy. We estimate a Markov transition matrix to shed some light on the dynamics of those variables. In section four, we introduce a dynamic multinomial logit model based on Honoré and Kyriazidou (2000) in order to be able to explain job and income dynamics. We detail the relation existing between this dynamic model and Markov transition matrices. We apply those results to compute individual probabilities of falling in unemployment for those having a job and probabilities of going back to work for those already unemployed. In section five, we model wage dynamics using the same econometric model, show that the underlying Markov transition matrix is not regular, so that the traditional POUM effect cannot exist automatically. In section six, we contrast preference for redistribution for three different groups (working, unemployed, not working) and show that the POUM effect, if it is present at the individual level, can largely be overridden by aversion to the risk of unemployment. Section seven concludes.

4.2 A simple model for redistribution preferences and income mobility

Let us consider a linear taxation-redistribution scheme relating the individual disposable income x_i to the individual gross wage w_i defined as follows:

$$x_i = (1-\tau)w_i + \tau\bar{w},$$

where $\bar{w}$ is the mean wage and τ the taxation rate. With this scheme, the government budget is in equilibrium. When $w_i < \bar{w}$, we have $x_i > w_i$. So individuals with a wage below the mean will be net receivers and consequently will vote for redistribution, provided that scheme is not going to be changed in the next future. This is the classical model of Meltzer and Richard (1981) designed for a static framework and based on the theory of the median voter of Romer (1975). In practice however, we often find that some individuals, despite being below the mean wage are against redistribution. This is the starting point of Benabou and Ok (2001c) and their POUM (prospect of upward mobility) effect. If individuals take into account their future income, they might anticipate that income mobility will make them better-off so that their predicted future income might be greater than the future mean income of the distribution. Benabou and Ok show that this effect does exist, provided we impose a mild restriction on the income mobility process. Consequently, studying preference for redistribution becomes just equivalent to analysing the properties of the income mobility process. See Alesina and Giuliano (2009) for a review.

Three main assumptions are made in Benabou and Ok that are:

1. *Individual incomes x_{it} are drawn from a common skewed distribution.*
2. *Income grows according to a continuous function f with a well defined expectation in x.*
3. *The function f is a concave non-affine function.*

As a consequence, there exists a current value of x, $x^* < \mathrm{E}_t(x)$ such that the individuals belonging to the income interval $[x^*, \mathrm{E}_t(x)]$ have a future expected income which is greater than $\mathrm{E}_{t+1}(x)$. With this simple consequence, as soon as individuals integrate their future income in their utility function, all those having an income greater than x^* will vote against redistribution, and not only those with an income greater than $\mathrm{E}_t(x)$, provided of course that they are not too much risk adverse.

There is a side assumption that is made in Benabou and Ok (2001c), but that does not seem to be used in their proof:

4. *Future income increases with current income in the sense of first order stochastic dominance.*

This last condition would be the equivalent of a Pareto assumption (everybody is better off with the future income distribution). This assumption is not innocent as it imposes a further restriction on the dynamics of the income mobility process, a regularity condition which entails the existence of a steady state (see the Appendix for more details).

We develop a small toy model around the lognormal distribution in order to illustrate the importance of each of the above assumptions when deriving the result of Benabou and Ok. We shall show that, when relaxing slightly some of them, notably the stochastic dominance assumption, their result is rather fragile. As a consequence, studying the income mobility process is certainly interesting per se, but will not exhaust the question of explaining individual preference for redistribution. A complete econometric model has to be build, explaining status mobility, income mobility and the entailed preference for redistribution for each group.

4.2.1 A lognormal example

Let us consider a population of n individuals that have an income which is log-normally distributed at time t with parameters μ_t and σ_t^2. That means:

$$x_{it} \sim \Lambda(\mu_t, \sigma_t^2), \qquad \log x_{it} \sim N(\mu_t, \sigma_t^2).$$

In order to exploit the properties of the lognormal process, we suppose that individual income grows according to an autoregressive process: [1]

$$\log x_{it} = \log a + b \log x_{i,t-1} + \epsilon_{it}, \tag{4.1}$$

where ϵ_{it} is a Gaussian white noise of zero mean and variance ω^2. The function f is thus defined as being:

$$f(x) = ax^b \exp(\epsilon).$$

This function is concave if $b < 1$. With a dynamics following (4.1), the income distribution in the next period will be also log normal, but with parameters $\log(a) + b\mu_t$ and $b^2\sigma_t^2 + \omega^2$ so that:

$$x_{i,t+1} \sim \Lambda(\log(a) + b\mu_t, b^2\sigma_t^2 + \omega^2).$$

That b lower than 1 constitutes a necessary but not a sufficient condition for the POUM effect to exist.

4.2.2 The POUM effect and risk aversion

Let us now introduce a particular utility function (Constant Relative Risk Aversion function) so as to be able to consider risk aversion:

$$U(x_i) = \frac{x_i^{1-\alpha}}{1-\alpha}, \qquad 0 < \alpha \leq 1.$$

The existence of a POUM effect requires that we can find individuals having an income at time t that is lower than the mean, but with a future expected

1. This is a variant of the Galton-Markov model extensively used for instance in Hart (1976) or Atkinson et al. (1992). See also Benabou and Ok (2001c) p. 475 for their income distribution and transition example.

income that has a greater utility than the utility of the mean of the next period income distribution. The expected utility of the predicted future income is computed as a conditional expectation:

$$\mathrm{E}_\epsilon U(x_{i,t+1}|x_{it}) = \int_\epsilon \frac{1}{1-\alpha}[ax_{it}^b \exp(\epsilon)]^{1-\alpha} dF_\epsilon.$$

Factorizing all the elements which are not a function of ϵ out of the integral, we get

$$\mathrm{E}_\epsilon U(x_{i,t+1}|x_{it}) = \frac{1}{1-\alpha}a^{1-\alpha}x_{it}^{(1-\alpha)b}\int_\epsilon [\exp((1-\alpha)\epsilon)]dF_\epsilon.$$

The integral then represents the expectation of a lognormal with parameters 0 and $(1-\alpha)^2\omega^2$ so that:

$$\mathrm{E}_\epsilon U(x_{i,t+1}|x_{it}) = \frac{1}{1-\alpha}a^{1-\alpha}x_{it}^{(1-\alpha)b}\exp(\frac{(1-\alpha)^2\omega^2}{2}).$$

This expected utility has to be greater than the utility of the mean of the future income distribution, namely $U(\mathrm{E}_\epsilon(x_{i,t+1}))$ which is equal to:

$$\begin{aligned} U(\mathrm{E}(x_{i,t+1})) &= U(\exp(\log a + b\mu + \frac{b^2\sigma^2+\omega^2}{2})), \qquad (4.2)\\ &= \frac{1}{1-\alpha}a^{1-\alpha}\exp((1-\alpha)b\mu_t + \frac{(1-\alpha)b^2\sigma_t^2+(1-\alpha)\omega^2}{2}). \end{aligned}$$

Equating these two expectations, we find the current value of income, x_{it}^* above which an individual will vote against redistribution:

$$x_{it}^* = \exp(\mu_t + \frac{b^2\sigma_t^2+\alpha\omega^2}{2b}).$$

A POUM effect is possible if in the current income distribution there exists $x_{it}^* < \mathrm{E}_t(X)$, which means:

$$\exp(\mu + \frac{b^2\sigma^2+\alpha\omega^2}{2b}) < \exp(\mu + \frac{\sigma^2}{2}),$$

the greater the distance $\mathrm{E}(X) - x_{it}^*$, the greater the POUM effect will be.

We have two interesting cases, depending on risk aversion:

1. When $\alpha = 0$ (risk neutrality), the POUM effect requires simply that $b < 1$. This is a simple concavity restriction on the transition function f which corresponds to the core assumption in Benabou and Ok.

2. In the case of risk aversion ($\alpha > 0$), the condition becomes more complex:
$$b(b-1) + \frac{\omega^2}{\sigma^2}\alpha < 0. \tag{4.3}$$
It includes a quadratic function of b and depends on the relative noise ratio ω^2/σ^2. When the noise in the income mobility tends to zero, we are back to the previous condition. For a strictly positive noise, more risk aversion implies more concavity and thus a lower b. This is true till a certain point because α has to be lower than a given number:
$$\alpha < \sigma^2/(4\omega^2). \tag{4.4}$$
Otherwise equation (4.3) has no solution. For a high degree of risk aversion, it is not possible to find a feasible value for b. So a large risk aversion can kill any possibility of a POUM effect.

The lower bound can be easily reached as can be seen from a rough calibration. A value of $\sigma^2 = 0.30$ corresponds to a Gini equal to 0.30, a most common value for gross income in Europe. [2] If $\omega^2 < 0.075$, then α has just to be lower than 1, which corresponds anyway to its upper bound. This value of ω^2 means a residual variance of 7.5% in regression (4.1) describing the mobility process. When $\omega^2 = \sigma^2$ meaning a much higher variance in the mobility process, then α cannot be greater than 0.25 to allow for a POUM effect. The parameter ω^2 represents uncertainty in the mobility process. For a small value of ω^2, society evolves at a regular pace and current income is mainly a function of past income. A larger value corresponds to higher social mobility which can be a higher risk of falling down in the social ladder due for instance to a greater risk of unemployment, or a greater chance of getting up. This is illustrated in Table 4.1. We note that with those calibrated values and

Table 4.1 – Percentage of a POUM effect with risk aversion

α	0.00	0.20	0.40	0.60	0.80	1.00
ω^2						
0.300	3.68	0.00	0.00	0.00	0.00	0.00
0.150	3.68	1.74	0.00	0.00	0.00	0.00
0.075	3.68	2.71	1.74	0.75	0.00	0.00
0.037	3.68	3.20	2.71	2.23	1.74	1.24

$b = 0.75$, the percentage of individuals with an income below the mean still

2. In the lognormal process, the Gini index is equal to $G = 2\Phi(\sigma/\sqrt{2}) - 1$. So that for a given value of G, we have $\sigma^2 = 2[\Phi^{-1}(\frac{G+1}{2})]^2$.

voting against redistribution is rather small. In an empirical application, risk aversion is going to be individual dependant, introducing thus heterogeneity.

4.2.3 Stochastic dominance and regular income dynamics

Stochastic dominance for the lognormal process was first analysed in Levy (1973). More precisely, his theorem 4 states:

Theorem 1. *Let F_1 and F_2 be two lognormal distributions with parameters μ_j and σ_j. F_2 first order dominates F_1 if $\mu_2 > \mu_1$ and $\sigma_1 = \sigma_2$.*

Stochastic dominance at the order one requires that the two processes have the same log variance. In our case, this condition implies that

$$\sigma_t^2 = b^2\sigma_t^2 + \omega^2 \Rightarrow b^2 = 1 - \omega^2/\sigma_t^2.$$

The second condition is a kind of growth condition. The requirement that $F_{t+1} > F_t$ entails the condition $\mu_{t+1} > \mu_t$, which translated in our example requires

$$\log a > (1-b)\mu_t.$$

The growth parameter a has first to be greater than 1 and secondly has to be an inverse function of b: the lower b, the greater a.

We are now in a position to interpret this condition of stochastic dominance. The first condition, which says that $\sigma_t^2 = \sigma_{t+1}^2$, guaranties stability for the process. In particular, if condition (4.4) on the risk aversion parameter is verified at the initial state of the system, it will be verified all the time. The proportion of individuals having an income between x^* and $\mathrm{E}_t(x)$ will remain constant and the upper bound on α will also remain constant. The absence of stochastic dominance at the first order can create a situation where for instance σ_t^2 decreases over time. That is a condition for Lorenz ordering. But at the same time, it becomes harder to meet the requirement on α when times elapses. So we can start from a situation where there is a POUM effect and that the POUM effect disappears after a certain time. Stochastic dominance at the order one eliminates irregular dynamic situations and thus might be an oversimplification when confronted to real data. It excludes for instance situation where inequality is decreasing if the log normal assumption is verified.

4.2.4 Heterogeneity

Feri (2012) in a similar lognormal model considers the case where the population is partitioned in two groups, a large group of unskilled individuals in proportion p with a low μ_1 experiencing a slow change of their income and a small group of skilled individuals in proportion $1-p$ with a higher μ_2 experiencing a quicker mobility. Consequently x_1^* will be greater than the mean income of the whole population $\mu = p\mu_1 + (1-p)\mu_2$ while $x_2^* < \mu$. The total effect will depend on the value of p and on mobility differences. In an empirical illustration on Italian data, Feri shows that the POUM effect is present when income mobility is treated as a whole and disappears when allowing for heterogeneity in income mobility. His result is obtained supposing a steady state and would be much more complex to derive outside this framework. This is thus another example where the POUM effect disappears.

4.2.5 Related empirical work

The POUM effect is easy to interpret. The income distribution has to evolve in such a way that it dynamically implements a redistributive scheme, so that no extra taxation-redistribution has to be implemented in order to reduce inequality. Provided that individuals correctly anticipate this mechanism and that they integrate future income in their utility function, they will vote against an extra redistributive scheme even if they are under the mean. However, we have shown that any departure from the four assumptions made in Benabou and Ok leads to situations where the POUM effect can disappear. Income mobility, heterogeneous risk aversion are complex processes that have to be examined in detail in order to relate them correctly to the preference for redistribution. We have thus to focus on the individual level. From the previous section, we have seen that individual preferences can be highly complex and non-linear even if the impact of each separate element is trackable. We have to study how different factors enter into the individual utility function.

There exists an important empirical literature containing evidence about the relation between preference for redistribution and income mobility, see e.g. Alesina and La Ferrara (2005) or Clark and d'Angelo (2008b) and the references quoted there. However, Clark and d'Angelo (2008b) focus on mobility between generations. Alesina and La Ferrara (2005) illustrates the importance of future income expectations which can dominate the current income effect. But they do not take into account heterogeneity and risk aversion. They measure income mobility by mean of an homogeneous Markov transition matrix and show that their result (the importance of upper income expectations on preference for redistribution) is robust to individual heterogeneity. In this paper, we shall point out that individual heterogeneity is of prime importance. Ravallion and Lokshin (2000), using Russian panel data insisted on heterogenous individual effects and on dynamics. But their panel survey contains information on preference for redistribution only for 1996. They concluded that differences in speed for predicted income mobility was a determinant factor.

4.3 Stylized facts from the BHPS

The British Household Panel Survey (BHPS) offers several advantages for studying income mobility and preference for redistribution. First of all, it allows modelling income dynamics as in Alesina and La Ferrara (2005) or Ravallion and Lokshin (2000). But it also contains variables on preference for redistribution as in Ravallion and Lokshin (2000), with however the advantage of covering many periods and not just a single one. The BHPS is based on a sample of British households that were first interviewed in 1991 (wave A). The members of these original households have since been followed and annually interviewed till year 2008 (wave R), which is the end of the panel. Income and demographic variables are present in all the waves while the specific variables reporting subjective opinions on redistribution are included only in waves B, D, F, H, K, M, P. We shall first detail the content of this survey for our main variables of interest and then explain our empirical strategy.

4.3.1 Job status

Individuals can be classified according to their job or physical status: working, unemployed, not working, students, retired, disabled. This is the variable `JBSTAT`.[3] We are going first to eliminate students, retired and disabled. Then we shall analyse a first modelling for the dynamics between the three job status: working, unemployed, non-participating. We assume that this is the key variable, to be considered first in a causal chain, because these categories are confronted to radically different mechanisms in term of exposition to risk and availability of opportunities in a context of globalization. We report in Table 4.2 the frequency of these three statuses over the whole period (1991-2008). On average, we have a total of 8 147 observations per year which a relatively small coefficient of variation due to the fact that the panel is not balanced. The main category is working. These three groups experi-

3. We consider both self employed and in paid employment as working in this paper.

Table 4.2 – Mean number of observations per year for each job status

	Working	Unemployed	Non-particip.	Total
Mean	6 767	461	919	8 147
Variation	0.237	0.183	0.203	0.219
Mean Frequency	0.827	0.058	0.114	1.000

These numbers are computed over the whole period considering waves A to R.

ence a totaly different mobility process as can be measured by estimating a Markov transition matrix as reported in Table 4.3. Within our panel period,

Table 4.3 – Average yearly transition matrix between job statuses

	Working	Unemployed	Non-particip.
Working	0.947	0.021	0.032
Unemployed	0.537	0.341	0.122
Non-particip.	0.291	0.037	0.672

This matrix was estimated by maximum likelihood using the variable JBSTAT from wave A to wave R. Prais index is 0.520.

94.7% of individuals working at time t stay working in the next period. Only 2.1% of them lose their job while 3.2% of them quit the labour market. This last group is mainly composed of females. Mobility within the unemployed group is much higher. 53.7% of the unemployed individuals find a job within one year while roughly 34.1% of them stay unemployed and 12.2% of them quit the labour market. The non-participating group also experiences a large mobility, but to a smaller extend. Roughly 29.1% of the members of that group returned to the labour market and find a job within one year while very few of them (3.7%) change their status to register as unemployed. So that finally, the overall mobility Prais (1955) index is equal to 0.520.[4] Mobility between these three groups is thus quite different and justify to treat them as separate for income dynamics and preference for redistribution.

4.3.2 Wage mobility

A conventional measure for social mobility consists in observing the evolution of annual household income (see e.g. Alesina and La Ferrara 2005). The key variable is `FIHHMN` in the BHPS, but it concerns only total household income of the previous month. `FIHHYR` is the yearly equivalent. In order to allow for individual heterogeneity, we have to distinguish between individual

4. index equals 1 means perfect immobility while 0 is perfect mobility. See the appendix for more details

income and household income, as anyway job status concerns only individuals. The key variable becomes the annual wage income reported by each individual. [5] This variable concerns the whole population. Those not working or unemployed have simply a zero wage. Those who have been partially employed have a positive wage because job status is recorded at the time of interview while earned wages are cumulated over the entire year.

Transition matrices have been widely used in the literature to model income mobility. They imply considering a discretization of the income variable. Benabou and Ok (2001c) make use of the estimated decile matrices reported in Hungerford (1993). Alesina and La Ferrara (2005) compute decile transition matrices using the PSID over 1972-1987 for household income mobility (and not individual income). They derive a measure of expected future income for each decile of the population using the algebra of Markov transition models. Other options are possible, for instance considering classes defined as fractions of the mean as in Jenkins (2000).

We have computed a quintile transition matrix using individual income over the complete period 1991-2008 and the same for household income. The results are reported in Table 4.4. These are average annual transition matrices. At the two extremes of the income distribution, there is a strong

Table 4.4 – Yearly income transition matrices

	Individual wages					Household income				
	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
Q1	0.735	0.183	0.049	0.023	0.009	0.673	0.208	0.065	0.033	0.021
Q2	0.120	0.652	0.175	0.041	0.012	0.178	0.502	0.221	0.071	0.029
Q3	0.027	0.135	0.645	0.168	0.024	0.066	0.178	0.478	0.224	0.054
Q4	0.013	0.032	0.131	0.690	0.135	0.037	0.073	0.179	0.515	0.196
Q5	0.008	0.013	0.023	0.107	0.849	0.025	0.037	0.063	0.165	0.709

These numbers were obtained by pooling data over the different waves A-R. Prais mobility index is 0.357 for individual wages and 0.531 for household income. Category 3 represent the class around the median.

probability of staying in the same category. Mobility is more important for household income than for individual wages. Those matrices can be used for prediction as they explain transition between states over time. Alesina and La Ferrara (2005) make use of them to compute a proxy for individual upward mobility. Category three is around the median. For each category, we compute the probability to move to categories four or five in the next

5. This refers to two income variables: net labour income for each employee and gross earnings for self-employed individuals. Note that the degree of non-response and under reporting tends to be higher for the self-employed group. For more details see for instance Jenkins (2000).

period or to stay in those categories. Alesina and La Ferrara (2005) allocate these probabilities to all individuals with an household income belonging to the same income quantile, assuming implicitly that these probabilities are not influenced by individual effects. Using the same figures, Table 4.5 shows that there is a large difference in mobility prospect for the first two quantile, depending which type of income is chosen. These transition matrices were

Table 4.5 – Probabilities to move to (or stay in) the last two income quintiles

Starting from	Q1	Q2	Q3	Q4	Q5
	Upper mobility			Upper stickiness	
Indiv. Wages	0.033	0.053	0.192	0.825	0.956
Household Inc.	0.054	0.099	0.278	0.710	0.874

computed under a Markov assumption, which is certainly hard to verify over the whole period. In the next section, we shall introduce individual effects in a more complex model.

4.3.3 Attitude to redistribution

It is now conventional to introduce in surveys questions with answers on an ordinal scale concerning subjective opinions on individual well-being, financial ease, satisfaction at work and many other topics. Studies on preference for redistribution like Ravallion and Lokshin (2000) for Russia, Alesina and La Ferrara (2005) for the US, Rainer and Siedler (2008) for Germany, Clark and d'Angelo (2008b) for the UK, to quote just a few, make use of this type of question concerning preference for redistribution.

In the International Social Survey Program (ISSP-2006), we find the following question *On the whole, do you think it should or should not be the government's responsibility to reduce income differences between the rich and the poor?* and a variant of it in the European Social Survey (ESS) *The Government should take measures to reduce differences in income levels.* In the British Household Panel Survey (BHPS) and in the Russian Longitudinal Monitoring Survey (RLMS), there is the restrict the rich question (RRQ): *Do you agree or disagree that the government must restrict the income of the rich?* These questions have been used by the above mentioned authors to document individual preference for redistribution. They correspond in fact to a kind of Robin Hood question: take money from the rich to implicitly redistribute it to the poor. These questions have the merit of clearly indicating who are the donors. It would be very difficult to formulate a precise question about taxation-redistribution as a real tax system is very difficult to describe. We should note however that those two questions (limit differences and restrict the rich) are phrased slightly differently so that they might produce different answers, at least answers that are not calibrated in the same way.

The BHPS contains the variable `OPPOLC`, reported on a numerical scale from 1 (strongly agree) to 5 (strongly disagree) corresponding to the following statement: *The government should place an upper limit on the amount of money that any one person can make.* This variable is not present in all waves but only in waves B, D, F, H, K, M, and P. There are 18,311 different individuals present in these 7 waves out of the 18 possible waves. The panel is unbalanced. The answered are ordered on a Likert scale (five points with indifference at the middle). It is clear from Table 4.6 that most individuals

Table 4.6 – preference for redistribution in the UK

	Agree				Disagree
	BHPS Restrict the Rich question				
Category	1	2	3	4	5
Percentage	4.40	17.40	16.20	47.20	14.80
Numbers	2377	9338	8706	25359	7953
	ISSP Government should reduce inequality				
Category	1	2	3	4	
Percentage	28.94	40.16	21.41	9.49	

The numbers corresponding the BHPS were obtained by pooling data over the seven waves B-P. The value 1 corresponds to strongly support, while 5 to strongly disagree. The ISSP percentages are those reported by Guillaud (2013) for 2006, where 1 corresponds to definitively should and 4 to definitively should not. The same ISSP question asked in Northern Ireland in 2007 gave a very similar pattern.

are mildly or strongly opposed to redistribution in the UK as reflection by the BHPS. But using a different question: *On the whole, do you think it should or should not be the government's responsibility to reduce income differences between the rich and the poor?* and the ISSP data as reported in Guillaud (2013) for the UK, we obtain quite a different majority because now 69.1% of the respondents think that at least it should be the government responsibility to reduce inequality. In the BHPS question, redistribution is implicit (if money is taken from the rich, it has to be redistributed somewhere), while in the ISSP redistribution is explicit as the question is clearly about inequality reduction and limiting the rich becomes implicit. Individuals use a different scale when answering the two questions.

The individual opinions reported in the BHPS evolve over time, probably under the influence of external shocks and personal history. A Markov transition matrix estimated between the five different states as reported in Table 4.7 show that those having a mild to a strong opposition to redistri-

Table 4.7 – Mobility among preference for redistribution using the BHPS Robin Hood question

	Agree	1	2	3	4	5	Disagree
Agree	1	0.315	0.348	0.116	0.163	0.058	
	2	0.094	0.406	0.201	0.266	0.033	
	3	0.031	0.178	0.362	0.379	0.049	
	4	0.015	0.088	0.133	0.630	0.134	
Disagree	5	0.012	0.038	0.050	0.384	0.515	

bution have rather stable opinions while the other could change more easily. A mobility Prais index is equal to 0.693, which is a rather high value. It is an empirical question to identify those who have a strong opinion against redistribution and who do not change and those who are much in favour of it, but apparently could change. We have access to this mobility in opinions because we are using a panel and not a cross section data set as in Alesina and La Ferrara (2005) for instance or a single year observation as in Ravallion and Lokshin (2000).

In Table 4.8, we have computed the average evolution of opinions, depending on the wage quantile of the responders. Rows sum to one as they model the distribution of a particular opinion over the income quantiles. The relation between income level and preference for redistribution is com-

Table 4.8 – Preference for redistribution as a function of wage quantiles

		$W = 0$	Q1	Q2	Q3	Q4	Q5
Agree	1	0.162	0.170	0.185	0.188	0.165	0.129
	2	0.187	0.198	0.194	0.175	0.145	0.102
	3	0.149	0.219	0.204	0.180	0.145	0.103
	4	0.127	0.161	0.164	0.173	0.188	0.187
Disagree	5	0.095	0.128	0.138	0.154	0.188	0.298

The first column indicates ordered answers to the Robin Hood question (1, 2,...). Each line indicates the proportion of answers as a function of income quantile. $W = 0$ indicates those having a zero wage, because they are not working or are unemployed.

plex. For those who strongly agree with redistribution, the occurrence of this opinion increases with wages till the median, and then decreases. For those who strongly disagree, the occurrence of this opinion strictly increases with wages. If we had used household income, the relation would have been strictly monotone. It is important to go into the household to measure differences of opinions.

4.3.4 Our empirical strategy

The theoretical model of section 2 has shown that a global POUM effect can disappear with a strong risk aversion, with individual heterogeneity, with irregular dynamics. In order to investigate the role of mobility upon redistribution preferences, we apply the following strategy shown in Figure 4.1 which describes a recursive causal chain which makes clear the role of individual heterogeneity, of income dynamics and of risk aversion (measured as fear of unemployment). We first explain job mobility between three main

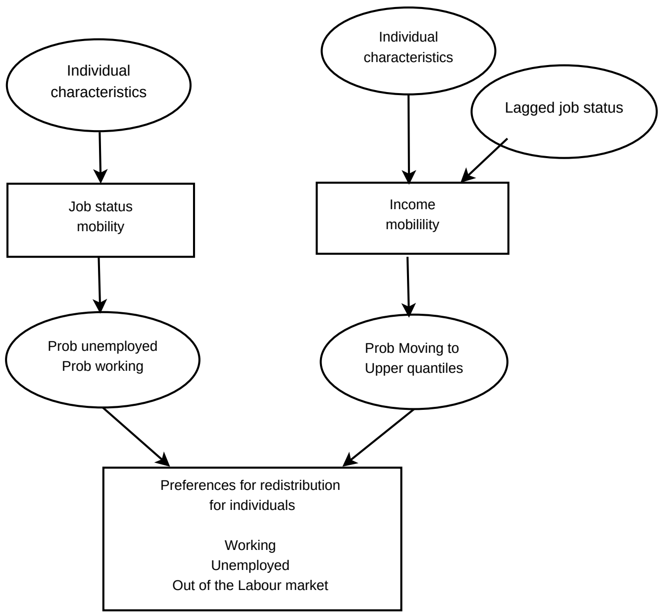


Figure 4.1 – A recursive empirical model to explain preference for redistribution

possible job statuses (employed, unemployed, not working) as a function of personal characteristics such as education, gender and age. With this model, we compute individual transition probabilities of being employed and of being unemployed in the next period which will be our main risk variables. At the same time, we explain individual income mobility as a function of individual characteristics and past job status. We derive from this second branch of the model the individual probability to move in the higher quantiles of the income distribution, which are above the mean. This is the key variable for modelling the POUM effect, taking into account individual characteristics. The third part of the model explains for each group (employed, unemployed, not working) preference for redistribution as a function some individual characteristics, the probability of future job status and the probability to have an income higher than the mean in the next periods.

Compared to the empirical strategy developed in the related literature detailed above, there are marked differences. The first difference comes from considering explicitly different job statuses as a main cause of variations in preferences for redistribution so that there cannot be a unique equation explaining those preferences. The probability of losing one's job is a measure of risk exposure and will be part of the model. This is not exactly risk aversion which measures how much utility is reduced by exposure to risk. But there is no satisfactory available measure of risk aversion in the BHPS. The variable `RISKA` (*Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?*) is present only in the last wave of the survey. The second difference comes from considering individual wage dynamics and not just household income dynamics as a proxy of the POUM effect as in Alesina and La Ferrara (2005). Individual wage mobility is certainly more directly linked to the dynamics of individual preferences and to the notion of effort. However, we shall add as an extra explanatory variable the dynamics of the remaining income of the household (the partner income mainly) and it will be apparent that this variable plays a different role, especially for those who have a zero income (unemployed or not working). The third and final difference come from taking into account the dynamics of opinions with a different persistence according to the current job status.

Note as a side remark and as a limitation to our study that our dependent variable on preference for redistribution appears only in 7 waves of the BHPS (roughly every 2 or 3 years) while jobs status and income dynamics are documented in all the 18 waves. We shall use the 18 waves to estimate the dynamic parameters and then use these parameters along with the variables contained in the 7 waves to compute the POUM proxy variable and the job status probabilities for estimating the final preference equations.

4.4 Modelling job status mobility

Our starting point is the existence of three main possible job statuses which can determine preference for redistribution because of a different exposure to risk and a different risk aversion: employed, unemployed and not working. Mobility between these three groups can depend on personal characteristics and on external shocks that we cannot model such as international competition, out-sourcing, well all the effects that globalization can have on employment. We shall model this mobility using a dynamic multinomial logit in order at least to filter the influence of personal variables. This model will allows us to compute individual transition probabilities of being employed and of being unemployed in the next period. These predicted variables corresponds to proxies for individual risk exposure and for individual opportunity.

4.4.1 A dynamic multinomial logit model

The job status transition matrix that we reported in Table 4.3 was estimated using a multinomial assumption, supposing that the sample was homogenous and that the mobility process was stationary. Each row of a transition matrix defines a multinomial process which is independent of the other rows (see for instance Anderson and Goodman 1957). To introduce observed heterogeneity, we have to consider a dynamic multinomial logit model which explains the probability that an individual i will be in state k when he was in state j in the previous period as a function of exogenous variables. [6] Using a simplified version of the model of Honoré and Kyriazidou (2000) and Egger et al. (2007), the unobserved propensity to select option k among K possibilities can be modelled as:

$$s^*_{kit} = x_{it}\beta_k + \sum_{j=1}^{K-1} \gamma_{jk} \mathbb{I}\{s_{i,t-1} = j\} + \epsilon_{kit}. \tag{4.5}$$

The observed choice s_{it} is made according to the following observational rule:

$$s_{it} = k \text{ if } s^*_{kit} = \max_l(s^*_{lit}),$$

with $s_{it} = 1$ if working, $s_{it} = 2$ if unemployed and $s_{it} = 3$ if non-participating. If the ϵ_{kit} are identically and independently distributed as a Type I extreme

6. Note that the log linear model of Tuma and Robins (1980) could also be a solution for explaining transition probabilities.

value distribution, then the probability that individual i is in state k at time t when he was in state j at time $t-1$ is given by:

$$p_{jk} = \Pr(s_{it} = k | s_{i,t-1} = j, x_{it}) = \frac{\exp(x_{it}\beta_k + \gamma_{jk})}{\sum_{l=1}^{K} \exp(x_{it}\beta_l + \gamma_{jl})}, \tag{4.6}$$

where x_{it} are explanatory the variables, α_k is a category specified constant common to all individuals and γ_{jk} is the coefficient on the lagged observed dependent variable attached to the transition between state j to state k. As the probabilities have to sum to 1, we must impose a normalization. We have chosen $\alpha_K = \gamma_K = 0$, $\beta_K = 0$. The standard estimation procedure is simple and can be done using a `R` package such as `VGAM`. However, execution time can be rather long, depending on the value of K and the sample size.

4.4.2 Initial conditions and individual effects

The original model of Honoré and Kyriazidou (2000) specifically introduces fixed individuals effects. Individuals effects are always difficult to identify in logit models. The usual practice of differencing the observations over the time dimension in order to eliminate the fixed effects is not possible in non-linear models. Honoré and Kyriazidou (2000) identify the fixed effects by selecting a small part of the sample verifying a precise sequence of four consecutive observations for the exogenous variables. And time dummies are not allowed. We cannot follow this approach first because we want to be able to compute characteristics for all the individuals and second because our sample is unbalanced. With an unbalanced sample the approach of Honoré and Kyriazidou (2000) would mean throwing away a great number of observations. So we have to rely on random individual effects. With random effects, the unobserved propensity to select option (4.5) is transformed into

$$s_{kit}^* = x_{it}\beta_k + \sum_{j=1}^{K-1} \gamma_{jk} \mathbb{I}\{s_{i,t-1} = j\} + c_{ik} + \epsilon_{kit}, \tag{4.7}$$

with a different random individual effect c_{ik} per option. It is no longer possible to use the `R` package `VGAM` because now the individual effects have to be integrated out:

$$p_{jk} = \int (p_{jk}|c) \times \phi(c|\Sigma)\, dc,$$

if we suppose that the random effects are jointly Gaussian of zero mean and variance-covariance matrix Σ (which can be diagonal). A second question arises which is not treated in Honoré and Kyriazidou (2000) and apparently elsewhere in the literature, the question of initial conditions in dynamic multinomial logit models. The way of treating the initial conditions in dynamic models with latent variables can be of prime importance as discussed in Xun and Lubrano (2014) for the dynamic Tobit model. The solution defended in Heckman (1981a) consists in modelling $s_{i0}|c_{ik}$ which implies finding the steady state of the model or an approximation to it. The solution advocated in Wooldridge (2005) is different as it models instead $c_{ik}|s_{i0}$ which is much simpler as it means for instance defining

$$c_{ik} = \alpha_{0k} + \alpha_{1k} s_{i0} + a_{ik}, \qquad a_{ik} \sim \mathrm{N}(0, \Sigma_a).$$

So, the latent variable is modelled as

$$s^*_{kit} = x_{it}\beta_k + \sum_{j=1}^{K-1} \gamma_{jk} \mathbb{1}\{s_{i,t-1} = j\} + \alpha_{0k} + \alpha_{1k} s_{i0} + a_{ik} + \epsilon_{kit}. \tag{4.8}$$

Compared to model (4.7), there is not much differences, we just have added extra regressors. The individual effects have to be integrated out using for instance the simulated maximum likelihood estimator detailed in Xun and Lubrano (2014) for the Tobit model. The base probability is

$$\Pr(s_{it} = k | s_{i,t-1} = j, x_{it}) = \frac{\exp(x_{it}\beta_k + \gamma_{jk} + \alpha_{0k} + \alpha_{1k} s_{i0} + a_{ik})}{\sum_{l=1}^{K} \exp(x_{it}\beta_l + \gamma_{jl} + \alpha_{0l} + \alpha_{1l} s_{i0} + a_{il})}. \tag{4.9}$$

However, due to huge sample size, the simulated maximum likelihood estimator is very time consuming. Assuming that the most important effect is provided by the initial conditions, we can constraint in a first step the variance of the random effects to be zero, so that `VGAM` can still be used in this particular case.

4.4.3 Marginal effects

The interpretation of the coefficients is rather complex in term of odd ratios, computed using the base probability formula (4.9). [7] It is simpler to

7. We have first that:

$$\frac{\Pr(s_{it} = k | s_{i,t-1} = j)}{\Pr(s_{it} = K | s_{i,t-1} = j)} = \exp(\alpha_k + x_{it}\beta_k) \exp(\gamma_{jk}).$$

We can then compute:

$$\frac{\Pr(s_{it} = k | s_{i,t-1} = K)}{\Pr(s_{it} = K | s_{i,t-1} = K)} = \exp(\alpha_k + x_{it}\beta_k) \exp(\gamma_{Kk}).$$

consider marginal effects which are defined as follows if x is a continuous variable:

$$\frac{\partial \Pr(s=k)}{\partial x} = \Pr(s=k)[\beta_k - \sum_l \Pr(s=l)\beta_l].$$

In the right hand side of this formula, $\Pr(s = k)$ is given by (4.9). This probability is a function of the vectors of exogenous variables and of random individual effects. As we need a single number for a marginal effect, we have to compute $\Pr(s = k)$ at the mean value of each exogenous variable and for the mean value of the random effect which is zero. When the marginal effect has to be computed for a dummy variable D (for example gender), the marginal effect is given by:

$$\frac{\partial \Pr(s=k|D,x)}{\partial D} = \Pr(s=k|D=1, x=\bar{x}) - \Pr(s=k|D=0, x=\bar{x}),$$

using again (4.9) and taking the random effects at their mean value.

4.4.4 Empirical job status mobility

The influence of dynamics and of the exogenous variables is best given using marginal effects as documented in the last two columns of Table 4.9. Even if they are strongly significant, initial conditions have a rather small effect in percentage. State dependance is very strong from employed to employed. Being employed in the previous period increases the chances to stay employed by 50% compared to starting from the baseline. And it decreases the chances of being unemployed by 5%. State dependence is much less important when starting from unemployed. Age has an U-shaped effect on the probability of being employed while it has an inverted U-shaped effect on the probability of being unemployed. Education has a positive effect on the probability of working and obviously a negative effect on the probability of being unemployed. But females have both lower probability of being employed or unemployed, which means that they mostly prefer to stay out of the labour market. We have chosen not to include wages because of possible endogeneity problems and because that variable is defined only for the working category.

As $\gamma_{Kk} = 0$, the ratio of the two above expressions is equal to: $\exp(\gamma_{jk})$,

$$\exp(\gamma_{jk}) = \frac{\Pr(s_{it}=k|s_{i,t-1}=j)}{\Pr(s_{it}=K|s_{i,t-1}=j)} \Bigg/ \frac{\Pr(s_{it}=k|s_{i,t-1}=K)}{\Pr(s_{it}=K|s_{i,t-1}=K)}\ , \tag{4.10}$$

which gives the interpretation of this coefficient. So $\exp(\gamma_{jk})$ refers to the ratio of the odds of being in status k compared to the baseline status K when having been in status j in the previous period over the same odds when having been in baseline status K in the previous period.

Table 4.9 – Estimation of a dynamic Multinomial Logit model for job status transitions using Wooldridge's initial conditions

			Random effects		Marginal effects	
Destination status	Work.	Unemp.	Work.	Unemp.	Work.	Unemp.
Intercept	3.625 (1.764)	18.223 (2.216)	3.488 (0.621)	18.106 (0.535)		
Origin: Work.	4.037 (0.033)	1.965 (0.064)	3.971 (0.043)	2.034 (0.069)	0.501	-0.054
Origin: Unemp.	1.770 (0.057)	2.865 (0.074)	1.877 (0.062)	2.662 (0.083)	0.143	0.217
Initial: Work.	0.837 (0.034)	0.208 (0.060)	1.271 (0.056)	0.034 (0.072)	0.050	-0.019
Initial: Unemp.	0.239 (0.057)	0.534 (0.073)	0.202 (0.065)	0.629 (0.091)	-0.003	0.015
$\log age$	-2.217 (0.995)	-10.229 (1.261)	-2.218 (0.352)	-10.258 (0.304)	0.155	-0.226
$(\log age)^2$	0.328 (0.139)	1.397 (0.178)	0.322 (0.052)	1.391 (0.046)	-0.019	0.030
mid educ	0.384 (0.035)	-0.139 (0.049)	0.407 (0.041)	-0.122 (0.058)	0.030	-0.018
high educ	0.641 (0.038)	-0.195 (0.055)	0.697 (0.044)	-0.141 (0.064)	0.044	-0.025
female	-2.051 (0.049)	-2.430 (0.057)	-2.052 (0.052)	-2.410 (0.064)	-0.052	-0.013
Random effect	0.0 (–)	0.0 (–)	0.577 (0.240)	1.053 (0.244)		
N. Obs.	115 982		115 982			
log-likelihood	-31 524		-31 406			

We used the routine `vglm` of the package `VGAM` in `R` to estimate this model without random effect, corresponding to columns 1 and 2. Standard errors in parentheses. Year dummies were included, but not reported. Unbalanced panel, waves A to R. Wooldridge initial conditions. Column 3 and 4 correspond to a model with random effects estimated using simulated MLE. Marginal effects are computed using the random effect model with initial conditions.

4.4.5 Unemployment risk, chances to go back to work

We shall now exploit the properties of model (4.9) is order to derive individual probabilities of being unemployed or being employed in the next period. These individual probabilities will have a strong potential influence on the POUM effect. The theoretical model has shown that a POUM effect can disappear if risk aversion is too large. With globalization, a major risk is now unemployment for those having a job. At least, for the least qualified individuals, delocalization and unemployment is felt as a major danger when they are working. On the other side, we might have a different effect for those already in an unemployment state, because they will have less to loose.

The probability of being in state k in the next period, conditionally on the fact that the individual is in state j in the previous period and on socio-demographic controls is given by (4.9), using the complete series of the covariates, but individual effects taken at their mean value. Using the estimated coefficients of Table 4.9, we calculate the exposure to the risk of being unemployed in the next period and the exposure to the chance of being employed in the next period for each individual. That allows us to build two new variables that we call $\Pr U$ and $\Pr W$. In Figure 4.2, we have plotted the sample

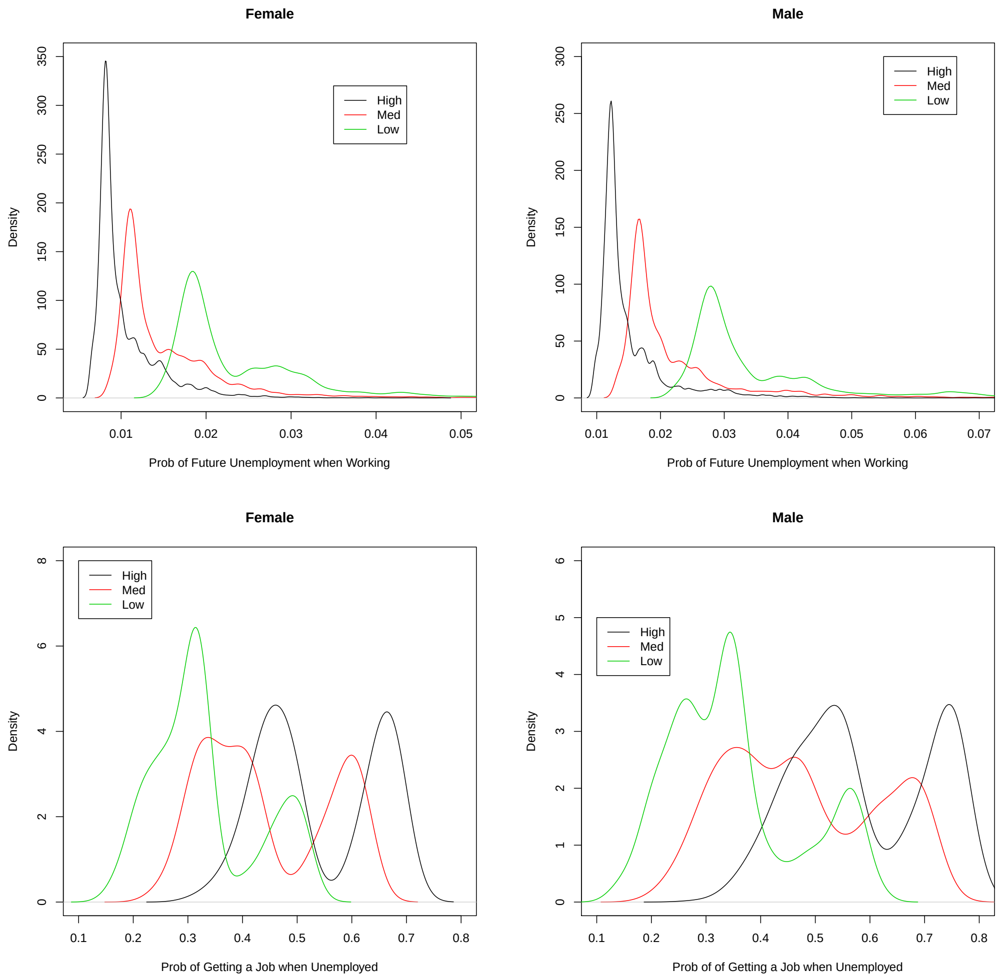


Figure 4.2 – Predicted job status probabilities

density of these predicted job status probabilities (conditioned on past job statuses and observed individual heterogeneity). As we have dummy variables included in (4.6), it is better to present these graphs for each category. We have separated males and females on one side and education levels on the other side. When working, males have more chances to fall in unemployment than females while their education level acts as a safety net. However, when unemployed, males have more chances to recover a job than females. The less educated individuals are, the more dispersed are their odds of changing their status. The density estimates of $\Pr W$ is multimodal because of the importance of initial conditions.

4.4.6 Implicit transition matrices for job status

The dynamic multinomial logit model can be seen as an alternative for estimating a Markov transition matrix. We can exploit the conditional probabilities given (4.9) to reconstruct the first $K-1$ lines of the transition matrix P and using the identification restrictions $\alpha_K = \gamma_K = 0$, $\beta_K = 0$ for the last line. The last column of the matrix is found using the constraint that each line sums up to 1. Of course, in order to obtain a single probability, we have to take the covariates at their sample mean as well as the random effects. As an illustration, we derived two transition matrices, one for males, one for females, computed at the mean value of the other exogenous variables. We report the results in Table 4.10. If the average of these two matrices look

Table 4.10 – Implicit conditional transition matrices

	Working	Unemployed	Non-particip.
		Males	
Working	0.963	0.019	0.017
Unemployed	0.605	0.290	0.103
Non-particip.	0.461	0.074	0.464
		Females	
Working	0.940	0.015	0.043
Unemployed	0.540	0.217	0.241
Non-particip.	0.266	0.035	0.697

much the same as the marginal one given in Table 4.10, there are huge differences between males and females for the unemployed and the not working lines. Males are almost always participating. Their most likely alternative is between working or being unemployed. Females mostly do not stay unemployed. They either go back to work or leave the labour market. When they have left the labour market, they have a strong tendency to stay in that state.

4.4.7 Testing for the validity of the model

When estimating a simple Markov transition matrix, several assumption have to be made as stated for instance in Shorrocks (1976). These assumptions are population homogeneity, Markov of order one and time homogeneity. At least two of these assumptions can be easily tested using a dynamic multinomial logit.

The first assumption that can be tested is the presence of observed individual heterogeneity. A LR test between a pure dynamic multinomial logit model with initial conditions (Log Lik. = -33 329) and the same model with exogenous variables, but no time dummies and no random effects (Log Lik. = -31 577) give a statistics of 3 504 with 10 degrees of freedom, so the pure dynamic model is rejected with a P value of 0.000. We had an example of individual heterogeneity with the two transition matrices for males and females which were clearly different. Finally, the significance of random individual effects is accepted using from the last line of Table 4.9 (LR statistics of 236 and 2 degrees of freedom).

The second assumption that can be tested is time homogeneity. This can be tested by comparing the dynamic model with observed individual effects (Log Lik. = -31 577) and the same model with time dummies (Log Lik. = -31 524) (one for each year). The LR test has a statistics of 106 with 32 degrees of freedom and again a P value of 0.000. The rejection of time homogeneity can be explained by the presence of a business cycle. The transition between working status and unemployed status highly depends on the economic activity in the short term, but is also depends on the effects of globalization in a longer term.

It would be quite difficult to test the Markov assumption, for instance to test the first order Markov against a second order Markov. That would mean estimating a model with k^2 possible states which would be very cumbersome.

So clearly, these test show that a simple Markov transition matrix would be rejected by the data. Covariates have to be taken into account because of population heterogeneity.

4.5 Wage mobility in an heterogeneous world

The possibility of a POUM effect at the aggregate level depends primarily on the properties of the income dynamics. In our simple theoretical model, the coefficient b has to be lower than 1. If income dynamics is modelled using a Markov transition matrix, the latter has to be first regular (each row stochastically dominates its predecessor which means that current income is an increasing function of past income) and then progressive in the sense that the expected income of the poor grows at a quicker pace than the expected income of the rich. See Benabou and Ok (2001c), Dardanoni (1993), Formby et al. (2004) for complementary details. There are two ways of estimating this matrix. By maximum likelihood as in (A.5), which means using no other information than past and current income. Or to use again a multinomial logit model, which is a convenient way of taking into account individual characteristics either observed or unobserved and past job status.

4.5.1 Wage dynamics and heterogenous Markov processes

Given a long panel (18 years), the same individual can experience spells inside each job status. For that individual, the wage variable can thus be zero at certain periods, can be slightly positive if the individual has been working only one part of the year or can represent a full year wage. We will thus have to explain a particular individual income dynamics that is certainly rather different from a household income dynamics, but which is more related to individual effort and individual circumstances, and certainly to the individual preference for redistribution.

Table 4.11 gives the estimation result for our multinomial model explaining transitions between wage quantiles, where a zero wage is taken as the baseline. In order to interpret those coefficients, we have to compute the marginal effects, evaluated at the average levels for all variables and assuming that the individual effects were taken at their zero mean. The marginal effects are reproduced in Table 4.12. Starting from Q_1 in $t-1$ increases the chances of moving in the next period one quantile ahead by 12%, but decreases the chances of moving to the two highest quantiles by 14% and 6%. When starting from Q_3 to Q_5, the chances of falling to Q_1 or Q_2 in the next period are decreased. When starting from Q_4 and Q_5, the chances of going to Q_3 or Q_4 are increased. The influence of age has an inverted U shape for staying in Q_1 to Q_3. It has a U shape for Q_4 and Q_5. Being educated decreases the chances to go to low quantiles (Q_1 to Q_2). But being a women decreases to chances to go to higher quantiles (Q_3 to Q_5). Having a job in the previous period decreases the chances to go down to the lowest quantiles, but increases the chances to go mainly to Q_3 and Q_4 by 33%. Being unemployed in the previous period has the reverse effect for higher quantiles (Q_3 to Q_5) by 15%. As all the regression coefficients are significant, we can suppose that the same is true for the marginal effects. These results demonstrate that the income dynamics is greatly influenced by individual characteristics so that it is hard to consider a unique homogeneous process valid at the macro level.

Table 4.11 – Wage quantile dynamics with observed and unobserved heterogeneity

	Q1	Q2	Q3	Q4	Q5
Intercept	5.643 (1.760)	16.084 (1.956)	1.960 (2.168)	−16.233 (2.459)	−43.255 (3.174)
Q1	1.672 (0.044)	1.295 (0.062)	0.671 (0.086)	0.087 (0.101)	0.051 (0.143)
Q2	1.346 (0.067)	3.546 (0.075)	3.043 (0.093)	1.839 (0.106)	1.432 (0.146)
Q3	1.055 (0.097)	3.038 (0.100)	5.038 (0.110)	3.949 (0.120)	3.004 (0.151)
Q4	0.999 (0.122)	2.399 (0.125)	4.132 (0.131)	5.787 (0.135)	4.859 (0.160)
Q5	0.961 (0.154)	2.012 (0.159)	3.109 (0.162)	4.462 (0.162)	6.805 (0.180)
Q1-ini-c	0.373 (0.043)	0.441 (0.051)	0.516 (0.059)	0.551 (0.067)	0.493 (0.085)
Q2-ini-c	0.382 (0.056)	0.794 (0.062)	0.878 (0.068)	0.647 (0.076)	0.520 (0.094)
Q3-ini-c	0.383 (0.069)	0.728 (0.075)	1.258 (0.079)	1.122 (0.085)	0.928 (0.098)
Q4-ini-c	0.258 (0.084)	0.459 (0.091)	0.930 (0.093)	1.283 (0.096)	1.305 (0.106)
Q5-ini-c	−0.030 (0.102)	0.009 (0.108)	0.277 (0.111)	0.554 (0.112)	1.289 (0.119)
Log age	−4.255 (1.003)	−10.675 (1.115)	−3.083 (1.232)	6.757 (1.389)	20.960 (1.769)
Log age2	0.528 (0.141)	1.315 (0.157)	0.202 (0.173)	−1.185 (0.194)	−3.143 (0.245)
EdHigh	0.301 (0.045)	0.547 (0.050)	0.894 (0.053)	1.429 (0.056)	2.071 (0.064)
EdMid	0.282 (0.039)	0.410 (0.044)	0.538 (0.047)	0.637 (0.051)	0.827 (0.061)
Gender	0.254 (0.045)	−0.102 (0.048)	−0.640 (0.050)	−0.951 (0.052)	−1.202 (0.056)
PJB = W	3.008 (0.047)	4.475 (0.070)	5.008 (0.095)	5.517 (0.119)	5.569 (0.152)
PJB = U	0.845 (0.051)	1.179 (0.082)	0.672 (0.120)	0.426 (0.156)	−0.501 (0.217)
Random effect	0.721 (0.220)	0.105 (0.407)	0.081 (0.308)	0.528 (0.196)	1.016 (0.141)
Log-likelihood:	-103 800.2				
$N \times T$:	115 982				

Zero wage is taken as the base line. Time dummies were added and are significant. Wooldridge initial conditions. Waves A to R, unbalanced panel. Standard errors in parentheses. Taking into account random effects improved the likelihood function from -103 949.5 to -103 800.2, but did not change significantly the estimation of the parameters. Year dummies were added, but not reported.

4.5.2 Implied transition matrices and individual heterogeneity

Let us now shed some more light on the Markov properties of the wage dynamics. The process is not time homogeneous as a log likelihood ratio test

Table 4.12 – Marginal effects for wage quantile dynamics with observed and unobserved heterogeneity

	Q1	Q2	Q3	Q4	Q5
Q1	0.240	0.017	-0.041	-0.049	-0.019
Q2	-0.261	0.409	0.091	-0.031	-0.017
Q3	-0.357	-0.029	0.514	0.080	-0.005
Q4	-0.374	-0.131	0.068	0.574	0.068
Q5	-0.374	-0.147	-0.037	0.104	0.659
Q1-ini-c	-0.016	-0.001	0.016	0.018	0.003
Q2-ini-c	-0.058	0.040	0.054	-0.002	-0.007
Q3-ini-c	-0.091	-0.034	0.106	0.047	0.004
Q4-ini-c	-0.094	-0.070	0.041	0.110	0.041
Q5-ini-c	-0.065	-0.066	-0.002	0.049	0.098
Log age	-0.522	-2.346	-0.408	1.601	1.625
Log age2	0.094	0.328	0.044	-0.236	-0.229
EdHigh	-0.103	-0.077	0.007	0.110	0.095
EdMid	-0.037	-0.012	0.021	0.029	0.018
Female	0.123	0.079	-0.055	-0.103	-0.058
PJB = W	-0.096	0.149	0.212	0.200	0.080
PJB = U	0.108	0.097	0.014	-0.001	-0.007

for the absence of time dummies in the above model has a value of 230 with 80 degrees of freedom, which makes a P-value of 0.000. So time homogeneity is rejected. Which means that income dynamics differs over the different years of the panel and that a single transition matrix cannot represent truthfully the income dynamics over the whole period. We could derive an average transition matrix using the estimation results of Table 4.11 and the mean values of the covariates, including the time dummies. But this would not be very meaningful, because the main interest of our model is to take into account individual heterogeneity. It is better to report the distribution of the implied individual probability of mobility which are directly related to the POUM effect.

4.5.3 Probability to go to the the upper quantiles

We computed the individual probabilities of moving to categories Q4 and Q5, using (4.9). From Figure 4.3, it is easy to see that males and females have very different income mobility opportunities. We have reduced income to wages. There are clearly two categories for males (those who can move to the upper categories and those who can stay in the upper categories)

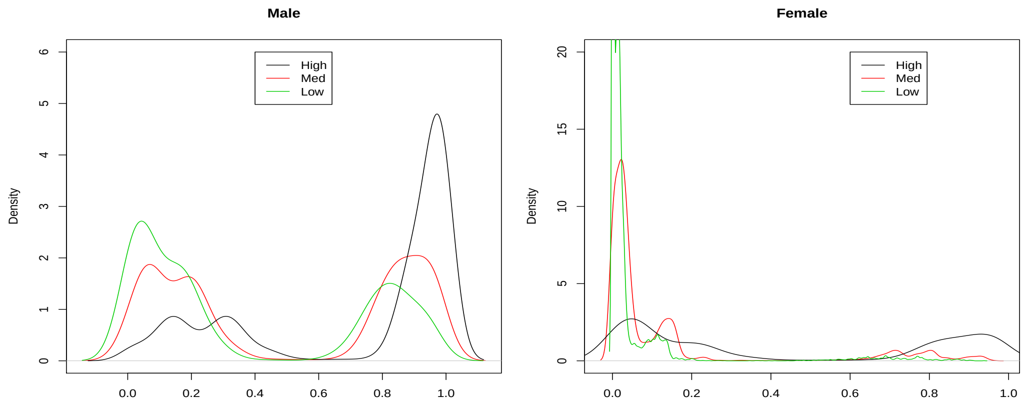


Figure 4.3 – Probability to move to or stay in the upper categories (Q4-Q5) as a function of education level for males and females

when there is mainly one for the females (those who can move to the upper categories). Both for males and females, education has a large impact for the probability of upper mobility. And that impact is much more important for males than for females. Low educated females have practically no chance of moving up when males do seem to have some chances. For the females, only those with a high education level could stay in the upper categories and this group is of a lower importance than the corresponding category for the males.

4.6 Preference for redistribution and income mobility dynamics

Opinions about redistribution, as approximated by the variable `OPPOLC`, are not constant over time, perhaps just because globalization is reshuffling income positions over time. Individual learn about their true position, experience shocks, so a dynamic model is essential to model their preference for redistribution. This is made possible by considering a panel, even if the variable `OPPOLC` is not present in all the waves, but only mostly every two years. We shall estimate for each group (working, unemployed, out of the market) a dynamic ordered probit model with unobserved individual random effects. Random individual effects and ordered probit are easily combined because we shall use the van Praag and Ferrer-i-Carbonell (2004, 2006) POLS version of the model. A key element of explanation will be the probability of being unemployed and the probability of getting a higher income and we managed to compute those probabilities for each individual. However, many individuals live in a family, so the mobility prospect of the other members in the household might also be important. Of course it has to be modelled in a different way. For each individual, we computed the income of the other members as the difference between individual wage income and household income. We then model the dynamics of this new variable that we call *Himw* using an unconditional quantile Markov transition matrix using five quantiles. We define the probability of having a higher complementary income *Himw* as the sum of the probabilities of the last two upper quantiles and attributed the adequate line probability to the current quantile of the household income.

Remark:

> Altruism can be a significant determinant for preference for redistribution. However, there is no altruism variable like (*Should we help the poor*) in the BHPS. There is only the `OPCHD` variable: *If you had to choose, which quality on this list would you pick as the most important for a child to learn to prepare him or her for life?* This variable is present only for waves C, D and E (1993, 1994 and 1995).

4.6.1 The working group

As we estimated a POLS model, the coefficients are directly interpretable. [8]

We find some similar results with the literature concerning socio-economic variables such as Females being more in favour of redistribution as well as individuals having children, when the marital status has no influence (see e.g. Alesina and La Ferrara 2005). But the similarities stop here, because we are considering three different groups and not the entire population. The main interest concerns of course the presence or absence of a POUM effect.

The POUM effect is present with two channels: the probability of having a wage higher than the 60% majority in the next period (`Prob Upper`)and the probability that the complementary household income will be higher than the 60% majority in the next period (`Prob Upper Himw`). Both variables have a strong effect. However, the wage level or the level of the complementary household income do not have any impact, as the tested restriction that their coefficient was zero has not been rejected. The effect of the income levels is dominated by the probability variables which are very different depending on

8. POLS means Probit OLS. It is based on a transformation of the endogenous variable into a normally distributed scaled variable so that the resulting model is the linear regression model. See van Praag and Ferrer-i-Carbonell (2004, Chap. 2) or Van Praag and Ferrer-i-Carbonell (2006).

Table 4.13 – Redistribution preference when in a working spell

Variable	Coef	Std. dev.	t-stat.
Intercept	-0.775	(0.647)	-1.198
1994	-0.042^{***}	(0.014)	-2.859
1996	-0.120^{***}	(0.015)	-7.698
1998	-0.139^{***}	(0.015)	-8.765
2003	-0.039^{***}	(0.015)	-2.613
2001	0.045^{***}	(0.015)	2.957
2006	0.074^{***}	(0.016)	4.504
y_{t-1}	0.098^{***}	(0.005)	17.449
Num. children	$-0.008^{\cdot}$	(0.005)	-1.608
Female	-0.180^{***}	(0.014)	-12.089
ln age centered	$0.607^{\cdot}$	(0.364)	1.666
ln age^2 centered	-0.104^{**}	(0.051)	-2.028
Prob Upper	0.277^{***}	(0.024)	11.569
Prob Upper$\times \mathbb{1}$(wageQ<4)	0.130^{**}	(0.062)	2.068
Prob Upper Himw	0.158^{***}	(0.017)	9.194
Prob U	-7.617^{***}	(0.838)	-9.086
Prob U$\times \mathbb{1}$(wageQ<4)	2.362^{***}	(0.773)	3.056
Better	0.060^{***}	(0.008)	6.943
Var(Intercept$\|i$)	0.343		
Var(Residual)	0.457	Intra Class = 42.9%	
Log-likelihood	-52269.4		
BIC	104752.1		
N	43075		
Individuals	15581		

The `marriage`, `wage` and `Himw` variables were present in the initial model. These overall restrictions were tested and accepted with a P-value of 0.273. $\mathbb{1}(wageQ < 4)$ is a dummy variable which is 1 when the individual belongs to the 60% lower part of the distribution.

the quantile to which the individual belongs. The POUM effect is greater for those who are in the lower categories as their total effect is $0.277 + 0.130 = 0.407$. Remember the particular distribution of the `Prob Upper` variable. The effect of expectations on personal wage is higher than the effect of expectation of complementary household income $(0.277 + 0.130 > 0.158)$. Finally the subjective variable `Better` contribute also to the POUM effect. All those variable play in the same direction of being against redistribution.

However, if the POUM effect is present, the magnitude of the effect of the probability of an income belonging to higher quantiles is totaly dominated by the effect of the probability of being unemployed in the next period. The risk exposure effect dominates the POUM effect when looking at the magnitude of the coefficients (even if we take into account the mean of variables). However, the effect of risk exposure is lower for those in the lower income quantiles (-7.617+2.362=-5.255). The greater sum an individual could lose, the greater is the effect of risk exposure on preference for redistribution. Consequently, the compensation hypothesis of Rodrik (1998) seems to be validated here. Following globalization, individuals ask for more protection against risk when they have a lot to lose.

Due to the configuration of the estimated parameters, the life cycle has a negative slope after age 20 which means that individuals become in favour of redistribution when aging. Finally, as we are in a dynamic model, we can qualify the speed of adjustment of opinions to changes in the situation of the individual. The mean lag of adjustment is very quick, of the order of two months, which is too quick when we compare this value to the conclusion of Piketty (1995) concerning dynasties.

Remark:

Ravallion and Lokshin (2000) have found, using the Russian panel survey, that income played a role in explaining preference for redistribution only for those who had an optimistic view concerning their future. This was shown by reporting group segmented regressions according to that variable (prospect for future income unchanged, increased or decreased), regressions which are relating preference for redistribution to the level of income. The impact of globalization concerns obviously the prospect for future income and differences in risk exposure. When individuals prefer the probability of a rise in their future income to a certainty of a redistributed income, they obviously prefer a gamble (on the future) to the sure thing which is redistribution. This is an immediate definition of risk aversion. At the level of our theoretical model of section 4.2 that would mean that we have at least two groups of individuals: a first group with a low value of α close to 0 and a second group with a much higher value for α (the risk aversion parameter). This would determine two different values for x^* for individual i in group j and thus two groups of voters with:

$$x^*_{ij} = \exp(\mu_t + \frac{b^2\sigma_t^2 + \alpha_j\omega^2}{2b}), \qquad j = \{1, 2\}.$$

If p is the proportion of voters in group 1, the proportion of individuals voting against redistribution while being poorer than the mean would be

$$F_t(\mu + \sigma^2/2) - pF_t(x^*_1) - (1-p)F_t(x^*_2).$$

That number would decrease when the proportion of individuals with a high risk aversion increases, for instance when unemployment is seen as a major risk for that group.

4.6.2 The unemployed group

Table 4.14 – Redistribution preference for unemployment spells

Variable	Coef	Std. dev.	t-stat.
Intercept	0.066***	(0.079)	0.828
1996	−0.102$^{\cdot}$	(0.053)	−1.916
2001	−0.097**	(0.042)	−2.296
y_{t-1}	0.321***	(0.027)	11.698
Female	−0.247***	(0.044)	−5.520
Prob Upper Himw	0.297***	(0.069)	4.259
Pr U	−0.575***	(0.121)	−4.726
Better	0.160***	(0.034)	4.674
Var(Intercept$\|i$)	0.081		
Var(Residual)	0.743	Intra Class = 10%	
Log-likelihood	−3 834.7		
BIC	7 749.1		
N	2 903		
Individuals	2 272		

Restrictions for year dummies (except 1996 and 2001), age, married, num. children, `Prob W` and household income variables were tested and accepted with P value of 0.108.

The behaviour of this group is rather different from that of the working group. Essentially, there is no longer any life cycle effect. The coefficient of inertia is much higher with 0.321, showing that this group may change its opinions in a much slower way than the working group and with a mean lag of one year. Moreover, the variance of the random individual effect is much lower, so that intra-class correlation is just 10%. Females are even more in favour of redistribution than in the working group.

There are however similarities with the working group. The probability of staying unemployed has a strong effect on the demand for redistribution. The POUM effect is still present, but using a different channel, because the wage variable is no longer present. We still have a POUM effect only through the income dynamics of the partner with the `Prob Upper Himw` variable. The optimistic variable `Better` has a stronger effect than in the working group. This is in accordance with Scheier and Carver (1992) who noted from clinical observations the importance of being optimistic in desperate situations. Note also Benabou and Tirole (2006) who illustrate in their model the *attitude of parents who pass on to their children a view of the world where effort ultimately pays off and everyone gets their just desserts.*

4.6.3 The non-participating group

Table 4.15 – Redistribution preference for out of the labour market spells

Variable	Coef	Std. dev.	t-stat.
Intercept	-5.111^{***}	(1.352)	-3.780
1996	-0.114^{**}	(0.032)	-3.496
1998	-0.137^{**}	(0.031)	-4.372
2001	-0.065^{**}	(0.025)	-2.537
y_{t-1}	0.196^{***}	(0.015)	12.276
Female	0.441^{**}	(0.083)	5.312
ln age	2.358^{***}	(0.764)	3.085
ln age2	-0.328^{***}	(0.106)	-3.078
Prob Upper Himw	0.241^{***}	(0.046)	5.146
better	$0.049^{\cdot}$	(0.026)	1.911
Prob W	1.339^{***}	(0.164)	8.146
Var(Intercept$\|i$)	0.210		
Var(Residual)	0.508	Intra Class = 44.1%	
Log-likelihood	$-7\,253.4$		
BIC	14 619.7		
N	5 924		
Individuals	3 137		

The restriction for year dummies 1994, 2003, 2006, children, married, and household income were tested and accepted with a P value of 0.441.

This is the group where most of the differences are located. First of all in this group females are strongly against redistribution, contrary to the two other groups. There is also a strong POUM effect concerning the income of the partner. If that income is going to increase, females (who are the 95% majority of this group) do not want that the future income of their partner could be taxed. The second important factor is the probability to go back to work which explains reluctance to redistribution. This is the only group where that variable is significant and has an important impact. There is also the strongest life cycle effect. Young people before 38 are against redistribution while after 38 they become favourable to redistribution. This life cycle effect dominate the optimistic variable that is roughly not significant here. The mean lag of adjustment is 6 months.

4.7 Conclusion

We have shown that the model of Benabou and Ok (2001c) relies on specific assumptions, that, when relaxed, could invalidate their main macro result: *with a left skewed income distribution, there can be a majority against redistribution.* Relaxing the homogeneity assumption (considering individual characteristics, different job statuses), we have shown that personal income dynamics can vary a lot, depending on individual characteristics, so that the resulting mobility matrices no longer verify the regularity (monotonicity) assumption. The regularity assumption is necessary to have a smooth income mobility compatible at the macro level with the POUM hypothesis. When considering three different groups of individuals according to their job status, we have shown that their preference for redistribution was totaly different and explained by different motivations. The working group is the one which resemble the most to the population considered by Benabou and Ok: the individual prospect of upward mobility does favour a reluctance to redistribution. But taking into account the employment risk greatly counterbalance this reluctance and validates the compensation hypothesis of Rodrik (1998). The individuals having experienced an unemployment spell or having left the labour market have very different individual preferences.

We have also managed to get at least partly into the household, decomposing household income into personal wage income and complementary household income. And the dynamics of these two incomes is not the same and has not the same influence on preference for redistribution. We could further investigate this aspect and identify completely the characteristics of the partner. Are there asymmetries between the partners, depending on gender or personal income? Which role is determinant, that of the male or of the female inside a couple?

Many other aspects would be interesting to investigate, in particular the influence of the wrong perception of the income distribution. For instance, Forsé and Parodi (2007), using the inequality program of the ISSP, report that in most countries individuals have a fairly correct perception of intermediate inequalities, but under-evaluate largely the extreme tails of the income distribution. When asked about their perceived position in the income distribution, individuals in most countries have a very important tendency to situate themselves around the mean, which can be interpreted as if they answered according to their reference group and not according to the complete income distribution. Poor people have a high tendency to overestimate their position in the income distribution or underestimate the dispersion of the distribution while rich people behave in the opposite way. However, this fact is less discussed in the literature when studying the effect of mobility upon redistribution preferences. If we assume that all individuals have a correct perception of the income distribution, we implicitly assume that people react to the comparison between the status of themselves and others, as documented in Boyce et al. (2010) who claim that the rank effect dominates the income effect. But a measure of the individual income perception is of course not easily obtained. This is certainly an opportunity for our future research agenda.

Several questions remains unsolved in this paper and deserve further research. We have aggregated the self-employed category with the employed category. These two categories can have a different behaviour because for some individuals being self-employed is just a transitory status before unemployment (see e.g. Bogenhold and Staber 1991 and also Gershon 2014). We have chosen in this version of the paper not to distinguish the self-employed, even if they certainly have a job status dynamics that would be different from the other categories. However, as underlined in Jenkins (2010), the income variable of this category is rather difficult to collect. Most of the time in the BHPS, the income variable for the self-employed is imputed, because either these individual do not want to report their income, do not know it precisely or under report it. Consequently, analysing the income dynamics of this category can be very hazardous and certainly need a great care and attention.

Our aim was to situate income mobility, preference for redistribution in a changing world, quoting often the effects of globalization. We have not taken in account these effect in a direct way, because that would have meant at least introducing sector variables, documenting job classification. The only way we allowed globalization to have specific effects on different categories of individuals was when introducing unobserved individual random effects. This was easy for the POLS models of the last section. This entailed a lot of computational effort for the multinomial logit models where the increase in computed time was enormous (two weeks for income dynamics).

Chapter 5

Preference for redistribution and poverty perception in China: Evidence from the 2006 CGSS

Acknowledgement

Data analyzed in this chapter were collected by the research project "China General Social Survey(CGSS)" sponsored by the China Social Science Fundation. This research project was carried out by Department of Sociology, Renmin University of China & Social Science Division, Hong Kong Science and Technology University, and directed by Dr. Li Lulu & Dr. Bian Yanjie. The author appreciate the assistance in providing data by the institutes and individuals aforementioned. The views expressed herein are the author's own.

5.1 Introduction: Inequality in China

After nearly two decades of economic stagnation, China started market-oriented reforms in December 1978. Being the world's fastest growing economy over the last 30 years, growth came together with an inequality increase and the latter became one of the most important issues. The income distribution changed dramatically since the mid-1980s. According to the China Statistical Yearbook, the over-all Gini index has grown from 0.35 in 1990 to over 0.45 in 2006. Using the Chinese household nutrition survey, Chen and Cowell (2013) claim that in the post-millennium area, climbing on the income ladder has become more difficult.

The social reforms and the fast economic growth have affected different groups of people in different ways. The rural/urban gap has increased. The

`Hukou` policy can be thought of being the cause of this widening gap. This state policy was adopted to limit mass migrations from the land to the cities, to ensure both economic and political social stability. The state favours urban residents and discriminates against rural residents in resource allocation, such as education, job vacancies, social benefits, health care, etc... (Afridi et al. 2012). Moreover, the redistribution scheme is decentralized in China as it mostly depends on the local economic level. This signifies that the subsistence allowances are much lower in rural areas than in urban areas (Houkai and Xiaoxia 2009). Thus being "rural" or being "urban" entails a huge discrepancy in terms of living standards and mostly in terms of opportunities. On the other hand, production activities in rural areas depend mostly on land farming although the labour-output ratio can vary a lot. Urban poor people have no access to means of production and they relies more on low paid job salaries and subsistence allowances. Rural individuals can decide to move to or to work in cities in order to have a relative higher income, but they will not necessarily be registered as "urban" officially. Although these rural labour forces contributed to a very important part in the economic development and urban modernization, the discrimination entailed by the `Hukou` system prevents them from having access to the benefits of the fruits of development in an equal way (Wong et al. 2007). The differences in rural and urban living and working circumstances and the isolation of rural people might lead to divergences in the perceptions of poverty between rural and urban people.

The urban group could be the beneficiary of growth in urban areas, but we also observe a significant group of the urban population who lives in a state of poverty (Fang et al. 2002). For example, in the early 90s, a great number of urban workers experienced the privatization of state owned enterprizes where they worked. Following this privatization, they became laid-off workers. This event consistently influenced their own life and that of their families. They dropped into the disadvantaged group.

The groups who benefited from most of these social and economic reforms are the government officers, businessmen, and people who have relations to them. They climbed up fast on the social ladder and distanced themselves from the pack by far. The increasing inequality becomes a potential risk for the stability of the society. The inter-regional inequality is also important. The east coastal provinces developed much faster than the interior provinces due to a series of preferential policies (Demurger et al. 2002). This divergence should contribute to the over-all inequality.

In the literature, we have plenty of evidence and discussions about inequality in China, see for instance, Kanbur and Zhang (1999), Khan and Riskin (1998), Khan and Riskin (2002), etc. Although, how people perceive inequality is less discussed in the literature. Some researchers have noticed the ten times increased number of mass protests in China from 1993 to 1995. This leads to the conclusion that Chinese ordinary people are angrier about the rising inequality (Tanner 2006). The Chinese government also notices this risk. During the term of office of the former leadership of Hu Jintao and Wen Jiabao, they put forward a slogan called "harmonious society" as well as series of policies to stabilize the society and to inhibit the dissatisfied. However, in a recent paper of Whyte et al. (2009), the author provides some evidence averse to these worries. By studying a national-wide inequality attitude survey, he concludes that the dissatisfaction for inequality has been overestimated in China. For example, the rural group is not more depressed by the inequality than the urban group as was assumed. According to this author, the rural group is more dissatisfied by procedural injustice rather than by distributive injustice. Moreover, unlike the urban group, the rural group has a limited perception of the real social ladder and people of that group tend to compare themselves to people who live in the same village. These conclusions come from the fact that rural individuals report the inequality within their community which is moderate. On the other hand, because the urban group has seen many upstart examples around them, they tend to be more depressed by inequality.

One question related to the inequality perceptions is the preference for redistribution, which is not a well developed topic in China. It also reflects people's inequality perception and adverts to how individuals perceive themselves as compared to others. Moreover, the topic of preference for redistribution is a natural experiment that bestows the possibility to survey many other topics, such as altruism and risk aversion. Lastly, does inequality perception influence the preference for redistribution? The correlation between preference for redistribution and inequality perception arises as an important issue in this study. Since the forming of the preference for redistribution is rather complex, many factors that determine the preference are unobservable, especially the value orientation and psychological traits. In Xu and Liu (2013), the authors show the importance of both social justice recognition and self-interest variables for explaining preference for redistribution. Unfortunately, this paper neglects the possibility of endogeny caused by the correlation between preference for redistribution and justice recognition.

The aim of this paper is to provide some evidence concerning the forming of preference for redistribution and poverty perceptions in China and the relation between them. The effect of `Hukou` policy upon preference and perceptions is also discussed. This paper is organized as follows. Section 2 reviews the literature about preference for redistribution and some important theories in this domain. Section 3 introduces the data base and discusses the choice of the potential determinants of preference for redistribution. Section 4 discusses one of the key variables in the preference for redistribution study, which is the occupation prestige scale and its readjustment for the Chinese society. The econometrics model is introduced and discussed in section 5. Empirical evidence are given in Section 6 along with economic interpretations. Section 7 concludes with some discussion.

5.2 Preference for redistribution and perception of poverty: a literature review

The literature about preferences for redistribution started with the static model of Meltzer and Richard (1981), based on the median voter theory of Romer (1975). It assumes that if the median income is lower than the mean income and if the government does nothing more than taxing the richer group (above the mean income level) and redistributing the taxes to the poor, then they will be a majority of population who will vote for a higher tax rate.

This main result of the model of Meltzer and Richard (1981) can break down if we considerer a dynamic framework where the voter introduces his future income in his utility function. If it is so, then people who earn an income lower than the mean level today are not necessarily interested in a redistribution policy if they anticipate that they would climb up over the mean level tomorrow. This idea was formalized with the *prospect of upward mobility* (POUM hypothesis) of Benabou and Ok (2001c). Within a dynamic framework, the median voter theory may no longer hold because individuals maximize their inter-temporal utility where their expected future income appears as an argument. They can vote against redistribution if their anticipation function is a concave function of their income.

The two models (static and dynamic) were tested on different data set. Even if it is now out of fashion, the static model of Meltzer and Richard (1981) was tested by Karabarbounis (2011) using the OECD SOCX data set over 14 countries. He found that more inequality was related to an increase in the demand for redistribution. However, people have become much more concerned the dynamic model. Using several data sets, economists and sociologists found proofs of the "POUM" effect in majoritarian democratic societies. For instance, Alesina and La Ferrara (2005) used the GSS (General Social Survey) and the PSID to relate income dynamics to preference for redistribution in the US. They found a POUM effect. Using the BHPS, Clark and d'Angelo (2008a) found a POUM effect when analyzing intergenerational mobility.

Economically, the "POUM" hypothesis relies on the specification of a particular individual dynamic utility function where only self interest is at work. However, preferences could be impacted by ideology, culture and family traditions and not simply by income levels or by income expectations as detailed in Piketty (1995) or in Benabou and Tirole (2006). See also Neustadt (2011) and Scheve and Stasavage (2006) for the effect of beliefs and religion.

Researchers have paid a lot of attention to the role of political ideology that generates differences between voters, differences based on issues such as equality, fairness, and the role of government, see Alesina and Glaeser (2004), Bean and Papadakis (1998), Feldman and Zaller (1992). Most of the discussions are around the relation between preference for redistribution and ideology as led around stylized facts. For example, Alesina and Glaeser (2004) try to explain why the EU society is more supportive of redistribution while the US society is much less supportive. They found that in US the majority tends to believe that poverty is generated by a lack of effort while the EU society tends to impute poverty to misfortune. The author points out that the link between preference for redistribution and beliefs about the nature of poverty relies on the `sense of justice`: "if you believe that luck (or inherited wealth) determines differences in income, you are more favorable to redistribution. If you believe that individuals' effort and individual's ability determine income, you are less favorable to redistribution" (Alesina and Angeletos 2005). In other words, it is a common sense that people should hold responsibilities for factors which are under their own control (i.e. lack of efforts) while they hold no responsibilities to external factors which are out of their control (i.e. circumstances), see for instance Rawls (1971) and Sen (1980) and Sen (1999).

Of course, differences in perception for the role of effort, luck/misfortune and preference for redistribution go back to long lasting historical and cultural differences between the two sides of the Atlantic. Alesina and Glaeser (2004) point out that the correlations are strong and *provocative*. This indicate that we cannot explain preference for redistribution as a function of poverty perception, but that these variables have to be explained simultaneously. At the individual level, many factors including one's life experiences, family background, psychological traits, social attitudes, ethics and the world outlook are usually unobservable. However, people who believe that luck determines success might still know the importance of effort. Alesina and Angeletos (2005) further point out that recognitions of luck and effort can be influenced by economic and political policies while redistribution policies are the revealed preference for redistribution of the majority in a democratic society. Fong (2001) emphasises that poverty perceptions are correlated with self-interest variables which leads to the question: what are the determinants of poverty perceptions? The causality of preference and perceptions is complex and simultaneous, while it is possible to quantitatively measure the correlations among these factors conditioned on exogenous variables.

In a recent paper of Xu and Liu (2013), authors enter self-interest variables and "key to success" variables jointly in the preference for redistribution equation. However, they assume that "key to sucess" variables are exogenous. Ignoring the simultaneity of the system would entail a serious endogenous problem. Although their results provide some evidence concerning the `sense of justice`, the problem is now how to determine the relation between preference for redistribution and the poverty perception in an efficient way. We shall provide a specific econometric model for that.

5.3 The Chinese General Social Survey

The Chinese General Social Survey (CGSS) is an annual or biannual repeated cross-section survey designed to collect individual opinions on social trends and the changing relationship between social structure and quality of life in China. CGSS is a sub-project of the International Social Survey Programme (ISSP). Following the structure of the famous general social survey (GSS), the CGSS provides multi-dimensional information on both socio-economic characteristics, attitudes and values on social issues. For the same reason as for the GSS, the respondents of each survey wave are randomly selected so that they cannot be supposed to be followed repeatedly so as to avoid selection bias and so as to ensure that the sample is representative of the whole population in each wave. The first wave was collected in 2003 and the last wave in 2010. The first wave provides very limited information while the social value part of the 2010 wave is not yet published. In this paper, we choose the 2006 wave because it contains the richest information available on social values. In this wave, 28 provinces are included, including Beijing, Shanghai, and some of the other most developed direct-controlled municipalities. This makes a total of 9 517 observations.

The 2006 CGSS is organized in four parts: The individual socio-demographics characteristics, the occupation status, the household components and status, and most importantly for us the subjective attitude variables.

5.3.1 Exogenous variables

Two types of explanatory variables are considered:

- *Socio-demographic variables:* region, gender, birth cohort, party membership, religion belief, material status, rural/urban status and years of education.

- *Individual socio-economic variables:* income, occupation prestige, occupation mobility with respect to that of the parents, and subjective expectation of household future socio-economic status.

There is a total of 28 provinces included in the data set. We regroup them into three regions:

- E.C. China (41.7%): East coast of China. The most developed provinces and the big cities of China (including Beijing, Shanghai and Shenzhen) and the three northeast provinces. This region has the most developed industries and the most developed third sector.
- C. China (26.5%): Central China. Less developed than the E.C, including the traditional agriculture provinces (Henan, Hunan, Hubei, etc).
- W. China (31.8%): The west of China, the least developed region.

The weighted sample proportions are given in parentheses. Descriptive statistics are given in Table 5.1.

The income variable includes all sources of individual income received in the year 2005 (currency unit: RMB). The summary table is shown in Table 5.2. There are 805 income missing observations and 1 003 observations with a zero income.

A subjective measure of the future upward mobility is also considered, the self-reported question: *How do you perceive your future household financial situation in three yeas ahead*, is it *better* (coded as 1) or not (coded as 0, includes *the same* and *worse*).

5.3.2 Dependent variables: Social values and opinions

In the attitude part of the survey, respondents are requested to report their opinions on a four-level scale tracing agreement to a given proposition (1 for totally disagree and 4 for totally agree). We have selected the three following questions:

1. Government should tax the rich more to help the poor.
2. Individuals are poor because society is not well functioning, especially because of misgoverning.
3. Individuals are poor because they are lazy

Descriptive statistics for social values and opinions are given in Table 5.3. A first glance at this Table gives us the impression that these are very similar questions. The distribution of preferences for redistribution is roughly the same as that of the *poor.misgov* variable. The *poor.lazy* variable is distributed just as the complementary distribution of the above two variables.

Table 5.1 – Socio-economic descriptive statistics using individual weights

Gender	
female	0.516
male	0.484
Birth cohorts	
-1959	0.162
1960-1979	0.462
1980-	0.376
Party membership	
member	0.152
mass	0.846
Religious beliefs	
Believer	0.137
atheist	0.846
Living in a couple	
yes	0.866
no	0.134
"Rural" in 2005	
yes	0.619
no	0.381
(Rural)"Migrant worker" in 2005	
yes	0.060
no	0.940
Being "New urban" (change within 10 years)	
yes	0.046
no	0.954
Years of education	
1st Qu.	6
Median	9
average	9.1
3rd Qu.	12

Table 5.2 – Income distribution

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Income distribution	0	2 000	5 000	8 824	11 300	250 000

Table 5.3 – Attitude and perceptions

	1	2	3	4
Redis. Pref.	0.030	0.173	0.528	0.275
Poor. misgov	0.025	0.206	0.588	0.181
Poor. lazy	0.201	0.467	0.265	0.078

This again confirms the warning of potential endogeneity problems. The ideology variables and preference for redistribution are jointly determined by unobserved factors such as value orientations and character traits.

5.4 Occupation and Social mobility

Clark and d'Angelo (2008a) have shown that the inter-generational prestige mobility between father and son have a significant effect upon the preference for redistribution. In that paper, job prestige is measured by the Hope-Goldthorpe Scale(HGS) [1] which transforms the occupation norms over a continuous scale. A precise measure of the occupation prestige should reflect the social ranking and the social class of individuals which are correlated to their preferences and perceptions.

In China, the `Hukou` system also contributes to the differences in occupations as well as to the occupation mobility. The migration from rural to urban regions is constrained which would limit the opportunities (discrimination) of "rural" people. Consequently, the `Hukou` system may have far-reaching influence upon preferences and perceptions through their life-cycle experiences and occupations.

In the CGSS, the individual's current job occupation and father's occupation are coded using the usual Erikson, Goldthorpe, and Portocarero (EGP) classification. This corresponds to:

1. Category I (40%): farm labor
2. Category II (26%): skilled/unskilled worker
3. Category III (12%): self employed

1. The HGS is an occupational index that reflects the job's reputation and classifies jobs according to their social desirability.

4. Category IV (10%): lower sales-service/routine non-manual
5. Category V (11%): higer/lower controllers

The EGP classes are ranked on the basis of two dimensions: Employee monitoring difficulties and human asset specificity (required on the job training), see for instance Edlund (2008). Both the HGS and EGP scales are designed for measurement purpose in western societies. It is not evident that the ordering entailed by the EGP classification is well adapted for China. In the following subsections, we will discuss the properties of inter-generational mobility in China and its consequences. Based on the following discussion, we will show that the EGP occupation prestige scale is not suitable for China. As a consequence, we have to build a proper measure of occupation prestige and the corresponding occupation mobility instead of using directly the EGP classification.

5.4.1 Inter-generational mobility via Markov chain

Like any other occupation variable, the EGP classification provides not only the occupation categories but also their corresponding prestige ranking. If the assumed prestige ranking does not adapt properly to China, it provides a misleading information for analyzing the preference for redistribution. The occupation mobility is one way to verify the validity of the assumed prestige ranking because the mobility monotonicity property holds if and only if the prestige ranking is monotone increasing.

As we have this classification both for the respondent and his/her father (current occupation or before retirement), we can model inter-generational mobility. Using these five ordered categories, we estimate a weighted Markov transition matrix to model an inter-generational transition matrix [2] which is reported in Table 5.4. We see that mobility is lower in lower rows and that the first row is the most sticky one. The first row corresponds to the mobility probabilities of individuals having a father in the farm labor category. Implicitly, this refers to the mobility of individuals who originally come from rural areas which indicates a significant policy barrier effect brought by the `Hukou` system. When his father was working in a farm, an individual has a probability of 0.54 to also work in a farm. For all the other categories, the

2. The transition matrix is estimated as:

$$P_{jk} = n *_{jk} /n*_{j.},$$

where P_{jk} refers to the probability of moving from origin category j to destination category k. $n*_{jk}$ is the weighted frequency of observations that move from j to k and $n*_{j.}$ refers to the weighted frequency of all the observations that origin from j.

Table 5.4 – Inter-generational mobility, Prais Index = 0.774

	EGP:1	EGP:2	EGP:3	EGP:4	EGP:5
EGP:1	0.54	0.23	0.09	0.07	0.07
EGP:2	0.09	0.44	0.08	0.23	0.15
EGP:3	0.17	0.19	0.31	0.25	0.08
EGP:4	0.06	0.28	0.14	0.32	0.19
EGP:5	0.11	0.25	0.13	0.21	0.29

probability for an individual to occupy the same type of job as his father is much lower.

Table 5.5 – A monotonicity test

	J=1	J=2	J=3	J=4
EGP:1				
EGP:2	41.45	14.74	15.88	6.99
EGP:3	**-3.15**	5.30	-1.53	-3.43
EGP:4	4.14	0.36	4.64	4.11
EGP:5	-2.67	-0.52	-0.05	3.82

In Table 5.5, we report the monotonicity test based on the definition of monotone matrices defined by Benabou and Ok (2001a). The Davidson and Duclos (2000) type dominance test is $T_{21} = -3.15$, thus monotonicity is rejected. For more details about definitons and calculations see Appendix A.6.1. This implies that the normal EGP occupation classification is not a suitable ranking system for Chinese society.

5.4.2 Occupation mobility

The previous mobility analysis tells us a story about the validity of the occupation prestige ranking. In this subsection, we use the stereotype ordered regression (SOR) model[3] to revise the ordering of the EGP scale. We want to explain the social class segmentation determined by the EGP scale by observable control variables in order to estimate the implicit ranking of the social occupations that individuals have in mind. The model is given as follows:

$$s_{ki}^* = \alpha_k + \phi_k * X_i\beta + \epsilon_{ki}. \tag{5.1}$$

3. The "stereotype ordered regression" (SOR) is reported for instance in Anderson (1984). See Hendrickx (2000) for an implementation in `Stata`.

The s^*_{ki} corresponds to the latent score (propensity) of category k for individual i while the α_k refers to the category specific intercepts. X_i is a set of observed variables that controls for the human capital (years of education), basic demographic variables (birth cohort, gender and rural/urban). Similarly to the unconditional Markov chain analysis of previous subsection, we also include the occupation category of the father in order to introduce inter-generational mobility. ϵ_{ki} is the error term which follows an extreme value distribution. The probability is then delivered through a logit type link function.

Unlike the linear part of the standard multinomial logit model, a SOR model constrains the category specific linear parameters β_k to be the same over all the categories. So $\beta_k = \beta$ whatever the value of k. However at the same time, a new multiplicative variable ϕ_k is introduced which relaxes in a way that restriction. It serves to measure the ordinal scale of the destination category ladder while this is not considered in the standard model. For identification reasons, we have to impose $\phi_1 = 0$ and $\phi_5 = 1$. To understand the new scaling metric parameter ϕ_k, we can write the log odds ratio of the two event probabilities $P(y_i = k)$ versus $P(y_i = k')$ as:

$$\log\left[\frac{P(y_i = k)}{P(y_i = k')}\right] = \alpha_k - \alpha_{k'} + (\phi_k - \phi_{k'})X_i\beta. \tag{5.2}$$

We shall see that the explanatory variable effects are measured by multiplying the category constant parameter β by an estimated category scaling metric ϕ_k. The higher the distance between ϕ_k and $\phi_{k'}$, the higher the magnitude of the effect given by X. In order to estimate the scaling parameter ϕ_k and the linear parameter β, an iterative method is used.[4] Table 5.6 reports the estimate of the intergenerational mobility (father's occupation versus current occupation of respondent). We shall see that the scaling metric ϕ_k is not monotone increasing. The scale of category III (skilled/unskilled worker) is higher than that of category II (self-employed). This result is comparable to the finding of Wu (2007). In western societies, category III is ranked higher than category II, see for instance Ganzeboom et al. (1989). One possible explanation given by Wu (2007) is that Chinese society (or more generally all communist societies) has a long tradition to inhibit private property. And eventually, becoming self-employed is easier than finding a stable job in the administration. Self-employment is less preferred in China and only concerns

4. The estimation procedure is as follow: first take the ϕ_k scaling metric as given and estimate β, then take the estimated β as fixed and estimate ϕ_k. The standard errors of ϕ_k are not identified while the standard errors of β are conditional on the given scaling metric ϕ_k. For more details see Hendrickx (2000). The estimation is achieved by the "mclgen" and "mclest" commands in the software `Stata`.

farm labor. The highest gap occurs between the first place (category I) to the second (category III) on the ladder, after that the differences are much smaller. This result shows an extreme low prestige of farm labor. Clearly, this occupation is found only in the "rural" group. The "SOR effect" reported in Table 5.6 are the estimates of β. We see that higher human capital is associated with higher probability of upward mobility. Being rural reduces the upward mobility probability while it could be weaken by the birth cohort effect (being born after 1980). Individuals are more likely to have a decent occupation if their fathers' occupation prestige is higher. We should also notice the stickiness of farm labor category. Lower categories have important negative effects upon the mobility for the next generation, which might be due to the policy barrier made by the `Hukou` system. From this analysis we see that the ordering of the EGP scale is not suitable for China. Henceforward, we avoid inserting directly the EGP category information. Instead, we use the transformed scaling metric parameter ϕ_k because it corrects the order of the occupation category with information of the scaling metric of each category. A dummy variable of inter-generational upward mobility is then coded as 1 if the following mobility event $\mathbb{1}\{s_{father} = j, s_{son} = k\}$ between two generations satisfies the condition that $\phi_k > \phi_j$.

5.5 Econometric modeling of preference for redistribution

We have three opinion variables that correspond to discrete observations which might be correlated. To each of these m opinion variables corresponds a level of unobserved utility z_m. This level, for every $m \in \{1, 2, 3\}$ is explained by a linear combination of exogenous variables X so that:

$$z_m = X'\beta_m + \epsilon_m.$$

The observation rule, relating the unobserved utility level z_m to the response variable Y_m is

$$Y_m = k \text{ if } \tau_{m,k-1} < z_m < \tau_{m,k}.$$

Y_m is the observed category vector which is reported in the survey, taking the ordered values from 1 to 4 in our case. $\tau_{m,k}$ is the threshold parameter that locates the boundaries for the discrete responses over the support of the continuous latent utility variable z_m.The error term ϵ_m is supposed to be normal of zero mean. For identification reasons, the three variances are set equal to 1. The three ordered probit models can be estimated separately, if

Table 5.6 – Inter-generational occupation mobility

the SOR model	
Scaling metrics	
EGP:1	0.0000
EGP:2	0.6201
EGP:3	0.5865
EGP:4	0.8481
EGP:5	1.0000
intercepts	
EGP:2	0.454 (0.137)
EGP:3	−0.490 (0.133)
EGP:4	−0.608 (0.184)
EGP:5	−1.129 (0.214)
SOR effect	
Father EGP:2:	0.321 (0.143)
Father EGP:3:	−0.506 (0.256)
Father EGP:4:	0.821 (0.246)
Father EGP:5:	1.109 (0.196)
Cohort 60-70	−0.328 (0.126)
Cohort post. 80	0.456 (0.134)
female	−0.357 (0.080)
yeduc	0.361 (0.015)
Rural	−4.153 (0.129)
$Pseudo-R^2$	0.234
N	8007

the three error terms are uncorrelated. In this case the probability of a basic event is equal to:

$$\Pr(Y_m = k) = \Phi(\tau_{m,k} - X'\beta_m) - \Phi(\tau_{m,k-1} - X'\beta_m).$$

If the m error terms are correlated, the basic event is much more complex and the three ordered probit models have to be estimated jointly. This model is related to the multivariate probit model (see e.g. Cappellari and Jenkins

2003). But here of course the dependent variables are ordered and not just binary.

5.5.1 A trivariate ordered probit model

As a starting point, let us consider the distribution of the error term:

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix} \sim N(0, \Sigma) \tag{5.3}$$

where we have 1's on the diagonal of the symmetric covariance matrix Σ. If the off-diagonal elements (ρ_{mn}) are all 0, then the model reduces to three independent ordered probit models. In order that Σ be positive definite symmetric, the elements of ρ in

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{pmatrix} \tag{5.4}$$

must verify some constraints, see next subsection.

Let us now consider the probability of the trivariate event ($Y_1 = j, Y_2 = k, Y_3 = l$). The evaluation of this probability requires the evaluation of a trivariate Gaussian CDF:

$$\Pr[Y_1 = j, Y_2 = k, Y_3 = l] = \int_{\tau_{1,j-1}-\hat{z}_1}^{\tau_{1,j}-\hat{z}_1} \int_{\tau_{2,k-1}-\hat{z}_2}^{\tau_{2,k}-\hat{z}_2} \int_{\tau_{3,l-1}-\hat{z}_3}^{\tau_{3,l}-\hat{z}_3} \phi_3(\epsilon_1, \epsilon_2, \epsilon_3, \rho)\mathbf{d}\epsilon_1\mathbf{d}\epsilon_2\mathbf{d}\epsilon_3,$$

where $\hat{z}_1$, $\hat{z}_2$ and $\hat{z}_3$ are the linear predictors $X'\hat{\beta}_m$ ($m = 1, 2, 3$), ϕ_3 is the PDF of a trivariate normal distribution and ρ represents the vector of all correlation parameters. There are good numerical methods for evaluating a bivariate normal CDF that are included in standard packages. But for higher dimensions, simulation methods are usually preferred. In our case, because of the truncation problem, the GHK (Geweke-Hajivassiliou-Keane) simulator seems to be a good candidate because the truncations could be directly simulated.

In order to apply the GHK simulator, let us rewrite the previous event probability as a product of conditional and marginal probabilities:

$$\begin{aligned}
&\Pr(Y_1 = j, Y_2 = k, Y_3 = l) = \\
&\Pr(Y_1 = j) \times \Pr(Y_2 = k|Y_1 = j) \times \Pr(Y_3 = l|Y_1 = j, Y_2 = k) = \\
&\qquad \Pr(\tau_{1,j-1} < \hat{z}_1 + \epsilon_1 < \tau_{1,j})\times \\
&\qquad \Pr(\tau_{2,k-1} < \hat{z}_2 + \epsilon_2 < \tau_{2,k}|\tau_{1,j-1} < \hat{z}_1 + \epsilon_1 < \tau_{1,j})\times \\
&\Pr(\tau_{3,l-1} < \hat{z}_3 + \epsilon_3 < \tau_{3,l}|\tau_{1,j-1} < \hat{z}_1 + \epsilon_1 < \tau_{1,j}, \tau_{2,k-1} < \hat{z}_2 + \epsilon_2 < \tau_{2,k}).
\end{aligned} \tag{5.5}$$

The difficulty comes from the fact that the ϵ_m are correlated. Let A be the lower triangular Cholesky decomposition of Σ such that $AA' = \Sigma$. Let us introduce three *iid* standard normal random variables η_m so that we can express the ϵ_m as a linear combination of the three independent η_m:

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}, \tag{5.6}$$

or in an expanded notation:

$$\begin{aligned}
\epsilon_1 &= a_{11}\eta_1, \\
\epsilon_2 &= a_{21}\eta_1 + a_{22}\eta_2, \\
\epsilon_3 &= a_{31}\eta_1 + a_{32}\eta_2 + a_{33}\eta_3.
\end{aligned}$$

Following this triangular system, we can decompose the joint probability (5.5) into the product of three conditional independent Gaussian probabilities. The first marginal probability is defined as:

$$\Pr(\tau_{1,j-1} < \hat{z}_1 + a_{11}\eta_1 < \tau_{1,j}) = \Phi[\frac{\tau_{1,j} - \hat{z}_1}{a_{11}}] - \Phi[\frac{\tau_{1,j-1} - \hat{z}_1}{a_{11}}], \tag{5.7}$$

and can be evaluated directly because $\Phi(.)$ is the CDF of η_1. The second conditional probability is:

$$\begin{aligned}
&\Pr(\tau_{2,k-1} < \hat{z}_2 + a_{21}\eta_1 + a_{22}\eta_2 < \tau_{2,k}|\tau_{1,j-1} < \hat{z}_1 + a_{11}\eta_1 < \tau_{1,j}) \\
&\qquad = \Phi[\frac{\tau_{2,k} - \hat{z}_2 - a_{21}\eta_1}{a_{22}}] - \Phi[\frac{\tau_{2,k-1} - \hat{z}_2 - a_{21}\eta_1}{a_{22}}],
\end{aligned}$$

where $\Phi(.)$ is the CDF of η_2. The third conditional probability is:

$$\begin{aligned}
&\Pr(\tau_{3,l-1} < \hat{z}_3 + a_{31}\eta_1 + a_{32}\eta_2 + a_{33}\eta_3 < \tau_{3,l} \\
&\qquad |\tau_{2,k-1} < \hat{z}_2 + a_{21}\eta_1 + a_{22}\eta_2 < \tau_{2,k}; \tau_{1,j-1} < \hat{z}_1 + a_{11}\eta_1 < \tau_{1,j}) \\
&\qquad = \Phi[\frac{\tau_{3,l} - \hat{z}_3 - a_{31}\eta_1 - a_{32}\eta_2}{a_{33}}] - \Phi[\frac{\tau_{3,l-1} - \hat{z}_3 - a_{31}\eta_1 - a_{32}\eta_2}{a_{33}}],
\end{aligned} \tag{5.8}$$

where $\Phi(.)$ is the CDF of η_3. The first marginal probability can be evaluated directly, using a standard numerical routine for Gaussian CDFs. The second probability is conditional on the distribution of η_1, which is unobserved. The idea of the GHK algorithm is to replace η_1 by a random draw from a truncated Gaussian distribution in order to evaluate the probability of a basic event and write the likelihood function. Of course, several draws have to be made as we shall detail below. Let us call η_1^{*r} the r^{th} draw of η_1 so that we have now:

$$\Phi[\frac{\tau_{2,k} - \hat{z}_2 - a_{21}\eta_1^{*r}}{a_{22}}] - \Phi[\frac{\tau_{2,k-1} - \hat{z}_2 - a_{21}\eta_1^{*r}}{a_{22}}], \tag{5.9}$$

where η_1^{*r} comes from a truncated standard normal density with lower and upper truncation points equal to $(\tau_{1,j-1} - \hat{z}_1)/a_{11}$ and $(\tau_{1,j} - \hat{z}_1)/a_{11}$ respectively. The third conditional probability includes two Gaussian random variables, the same η_1 as before and η_2. We use the same η_1^{*r} as before and draw η_2 from a truncated Gaussian so as to have:

$$\Phi[\frac{\tau_{3,l} - \hat{z}_3 - a_{31}\eta_1^{*r} - a_{32}\eta_2^{*r}}{a_{33}}] - \Phi[\frac{\tau_{3,l-1} - \hat{z}_3 - a_{31}\eta_1^{*r} - a_{32}\eta_2^{*r}}{a_{33}}]. \tag{5.10}$$

This time, η_2^{*r} is drawn from a standard normal density with lower and upper truncation points $(\tau_{2,k-1} - \hat{z}_2 - a_{21}\eta_1^{*r})/a_{22}$ and $(\tau_{2,k} - \hat{z}_2 - a_{21}\eta_1^{*r})/a_{22}$. We explain in Appendix A.6.3 how to draw truncated random numbers using GHK algorithm.

Since the computation of Equation (5.7) is straightforward, we shall initialize the algorithm by computing it first and then recursively evaluating Equation (5.9) and (5.10). Now if we have R draws of η_1^* and η_2^*, the simulated probability is then the arithmetic mean of each probability given the r^{th} random draw of ξ^r (see Appendix C):

$$\overline{\Pr}(Y_1 = j, Y_2 = k, Y_3 = l)_{GHK} = \frac{1}{R}\sum_{r=1}^{R}[\Pr_1 \times \Pr_2^r \times \Pr_3^r]$$

where $\Pr_2^r$, $\Pr_3^r$ refer to Equations (5.9) and (5.10) respectively given r^{th} draw of ξ. Finally, the simulated likelihood function is given by:

$$L_{GHK} = \prod_{i=1}^{N} \overline{\Pr}(y_{i,m} = k)_{GHK}$$

for $m = \{1, 2, 3\}$ and $k = \{1, 2, 3, 4\}$ and $Y_m = \{y_{1m}, \cdots, y_{im}, \cdots, y_{Nm}\}$. The weighted likelihood function is:

$$WL_{GHK} = \prod_{i=1}^{N} \overline{\Pr}(y_{i,m} = k)_{GHK}^{w_i}, \tag{5.11}$$

where w_i is the weight value assigned to individual i as our data set is a weighted sample.

5.5.2 Monte Carlo simulation

Now let's consider a Monte Carlo simulation example in order to verify that our method is working correctly. We have selected a sample size of 1 000 and a number of replications equal to 1 000. We first draw the three independent explanatory variables X_1, X_2 and X_3 from a standard normal distribution with mean zero and standard deviation 1.5. Once we have the X, we select values for the βs so as to generate the latent utilities. We have selected the following structure:

$$\begin{aligned} z_1 &= 0.3 * X_1 - 0.6 * X_2 + 0.9 * X_3 + \epsilon_1, \\ z_2 &= 0.2 * X_1 - 0.3 * X_2 + 0.6 * X_3 + \epsilon_2, \\ z_3 &= -0.2 * X_1 + 0.9 * X_2 + 1.5 * X_3 + \epsilon_3. \end{aligned}$$

The coefficients of X in each equation are chosen arbitrarily. The error terms ϵ_1, ϵ_2, ϵ_3 are simulated from a trivariate normal distribution with zero mean and covariance matrix:

$$\mathbf{Cov}\begin{pmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\end{pmatrix} = \begin{pmatrix} 1 & 0.25 & -0.4 \\ 0.25 & 1 & 0.6 \\ -0.4 & 0.6 & 1 \end{pmatrix}. \tag{5.12}$$

The threshold parameters are chosen so as to correspond to the (0.25, 0.50, 0.75) quantiles of the $\hat{z}_m$. The ordinal responses $Y_{\{1,2,3\}}$ are then generated accordingly.

We report in Table 5.7 the mean bias when Σ is treated as an identity matrix, and the mean bias when Σ is estimated using the GHK simulator. Table 5.8 reports the same results for the MSE. It is well apparent from these tables, first that when there is correlation, a serious error is committed if the model is treated as three independent ordered probit models. And second, which is in a way the most important result, that our method using the GHK simulator managed to give quite accurate results as well as the access to the correlations.

5.5.3 Evaluation strategy

The trivariate ordered probit model was programmed using the software **R**.[5] The basic idea is to maximize the simulated likelihood function, this is

5. The **R** code of the Monte Carlo experiment is available upon request. Contact: decxun@hotmail.com

Table 5.7 – Mean bias for comparing three independent ordered probit with a trivariate ordered probit

	Independent ordered probits			Trivariate ordered probit		
	y_1	y_2	y_3	y_1	y_2	y_3
X_1	-0.219	-0.138	0.154	-0.000	0.000	-0.000
X_2	0.443	0.210	-0.697	0.002	0.001	0.003
X_3	-0.665	-0.419	-1.167	-0.004	-0.001	0.003
ρ_{21}		0.250			0.000	
ρ_{31}		-0.400			-0.005	
ρ_{32}		0.600			0.006	

Table 5.8 – MSE for comparing three independent ordered probit with a trivariate ordered probit

	Independent ordered probits			Trivariate ordered probit		
	y_1	y_2	y_3	y_1	y_2	y_3
X_1	0.050	0.021	0.026	0.0007	0.0006	0.0007
X_2	0.199	0.046	0.492	0.0010	0.0007	0.0023
X_3	0.447	0.178	1.374	0.0014	0.0009	0.0050
ρ_{21}		0.063			0.0016	
ρ_{31}		0.160			0.0018	
ρ_{32}		0.360			0.0015	

done via the package “maxLik” using the "BHHH" algorithm in **R**. In order to initialize the evaluation, a reasonable set of starting values has to be provided. In this study, the staring values are chosen from the coefficient estimated from the independent ordered probit models while the starting values for the correlations are set equal to zero. As we have discussed already, several constraints have to be imposed to ensure that the model is identifiable.

Another thing which is important in the evaluation of the model is how to ensure the positive definite property of the variance covariance matrix Σ. If ρ_{21} and ρ_{31} are freely chosen between 0 and 1, then the third term ρ_{32} must verify the constraint:

$$\rho_{21}\rho_{31} - \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2} \leq \rho_{32} \leq \rho_{21}\rho_{31} + \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2}. \quad (5.13)$$

A proof of this result is given in Appendix A.6.2. This condition is essential to improve the efficiency of the evaluation along with the iterations of the MLE process, otherwise an Acceptance-rejection sampling strategy should be used instead which is quite slow especially when the correlation dimension is high. Unfortunately, this condition does not seem to be commonly applied in most of the statistical softwares and packages. This condition could be generalized for a correlation matrix of higher dimensions.

5.6 Preference for redistribution in China

Table 5.9 reports the estimation of the trivariate ordered probit model (5.11) discussed in the previous section which explains answers to the three ordinal variables related to preference for redistribution and poverty perceptions. We present the unestricted version of the model. If we set all the individually insignificant coefficients to zero (lower than a 90% significance level), the log-likelihood value drops from -17 904 to -17 908.92, with DF of 20, so the joint null hypothesis is accepted. If we now try to restrict to zero the structural variances (equivalent to considering three independent ordered probit models), the likelihood value drops to -17189.73. The difference is 95.76 with DF of 3, so that the null restriction is rejected. The joint model is validated.

5.6.1 Structural correlation

Our model provides an efficient estimate of the correlation matrix among ordinal variables conditionally on the exogenous variables. Our three ordinal variables have to be correlated (mutually endogenous) for our trivariate ordered probit model to be justified. From Table 5.9, this is the case. The correlation is not important, but it is highly significant. More precisely, the correlation is very significant between the *preference for redistribution* variables and the two other variables. However, it is not significant between the "poor-misgovernment" and the "poor-lazy" variables. This means that these two variables provide independent information on "preference for redistribution" and are not negatively correlated. Remember that Alesina and Glaeser (2004), when comparing subjective poverty perception in Europe and in the US, were considering two exclusive justifications: lack of effort in the US and absence of luck in Europe. In China the two types of explanation can play a complementary role, and that at the same time in the same mind.

The correlation between *preference for redistribution* and *poor-misgovernment* is positive. This means that the variables explaining the opinion about mis-governance will have also an indirect influence upon the preference for redistribution with presumably the same sign. This means also that individuals thinking that mis-governance is a cause for poverty would tend to believe

Table 5.9 – Preference for redistribution and poverty perception

	Redis. Pref.	Poor-misgov	Poor-lazy
C. China	0.009 (0.033)	-0.090^{**} (0.033)	0.068^{*} (0.032)
W. China	0.005 (0.029)	-0.107^{***} (0.029)	0.151^{***} (0.029)
birth 60-79	0.085^{*} (0.043)	0.120^{**} (0.042)	-0.090^{*} (0.043)
birth post 80	0.117^{*} (0.046)	0.117^{**} (0.045)	-0.124^{**} (0.045)
female	0.069^{*} (0.033)	-0.035 (0.035)	-0.090^{**} (0.033)
party	-0.039 (0.036)	-0.089^{**} (0.034)	-0.054 (0.035)
believer	-0.088^{*} (0.041)	-0.158^{***} (0.040)	-0.106^{**} (0.037)
coupled	-0.006 (0.039)	-0.043 (0.040)	0.031 (0.038)
Rural	$-0.078^{\cdot}$ (0.041)	-0.092^{*} (0.039)	0.127^{**} (0.039)
yeduc	-0.006 (0.005)	-0.018^{***} (0.004)	0.003 (0.004)
ln income	0.100^{***} (0.020)	0.075^{***} (0.019)	-0.008 (0.019)
ln income squared	-0.009^{***} (0.002)	-0.007^{***} (0.002)	0.001 (0.002)
Occup. prestige	0.093 (0.069)	-0.021 (0.069)	0.030 (0.065)
Upward (fath./son)	-0.124^{**} (0.047)	0.018 (0.048)	-0.122^{**} (0.045)
Upward (moth./dau.)	-0.129^{*} (0.052)	$0.097^{\cdot}$ (0.054)	-0.030 (0.050)
better finance	0.001 (0.026)	0.027 (0.026)	0.163^{***} (0.026)
1|2	-1.787^{***} (0.093)	-2.157^{***} (0.091)	-0.689^{***} (0.089)
(2|3)-(1|2)	1.075^{***} (0.029)	1.263^{***} (0.032)	1.265^{***} (0.017)
(3|4)-(2|3)	1.491^{***} (0.018)	1.690^{***} (0.019)	1.001^{***} (0.019)
$\rho_{R.P,misgov}$		0.202^{***} (0.016)	
$\rho_{R.P,lazy}$		-0.086^{***} (0.016)	
$\rho_{misgov,lazy}$		0.013 (0.016)	
N		5417	
Loglik		-17 094	
R		25	

The two Upward dummy variables are 1 if the prestige of a son is higher than his father's, the same for the variable measuring the upward mobility of a daughter compared to her mother. Their are built using the revised order of occupation categories provided by the SOR model.
p-value codes: "***" for 0.001, "**" for 0.01, "*" for 0.05 and "." for 0.1.

that poverty has to be compensated by redistribution.

The correlation between *preference for redistribution* and *poor-lazy* is negative and significant. However, the correlation is much lower. So individuals thinking that the main cause of poverty is laziness are also less in favour of redistribution. However, because of the smaller correlation, that effect is less important than the previous one. Consequently the `sense of justice` detailed and explained in Alesina and Glaeser (2004) is also present in China as the impact of *poor-misgovernment* is more important than that of the *poor-lazy* for explaining the preference for redistribution.

The final consequence of the significant correlations is that even if a variable is not significant in the equation explaining preference for redistribution, it can have an indirect effect provided it appears significantly in one of the other equations.

5.6.2 Poverty perceptions

The estimates of the two equations corresponding to beliefs in the causes of poverty are reported in column 2 and 3 of Table 5.9. For these two equations, both regions and birth cohorts are significant. The estimated sign of these two sets of dummy variables in these two equations shows an exclusive pattern. Generally speaking, individuals living in the Central or in the West part of China (comparing to the East Coast region) support less the idea that poverty is generated by misgovernment while they tends to support the idea that laziness is the main cause of poverty. New generations support more the idea that poverty is caused by misgovernment rather than by laziness. The gender dummy variable enters as a significant factor only in the *poor-lazy* equation. Females tend to impute poverty to laziness more than males. The negative sign of the party membership dummy variable is expected in the *poor-misgov* equation. Party members tend to be more confident in the ability of the Party to fight against poverty. While being a member of the Party makes no significant differences in answering the *poor-lazy* question. The strong effect of party membership in the *poor-lazy* equation is then an evidence that a self-interest variable can influence poverty perception.

Having a religion belief has a negative and a significant effect in both poverty-origin equations. This might be because those individuals impute poverty to other external factors which are not surveyed in the data set. Living in a couple has no effect in both cases. Having more years of education decreases the degree of recognition that misgovernment causes poverty while it has no effect in the *poor-lazy* equation. The effect of log-income has an inverted U-shape. But as the first 25% quantile level of log-income (7.601) locates on the right hand of the curve peak, the effect of log-income has then only a monotone decreasing trend with an increasing speed. Occupation prestige has no effect in both equations. Having an upward mobility experience (compared to the father) drifts negatively the recognition that idleness causes poverty. Lastly, people who anticipate an upward household financial situation agree more that laziness causes poverty.

Rural individuals tend to impute poverty less to misgovernment while they impute it more to laziness. This result might not be coherent to the evidence found in section 5.4.2. Farm labour (rural group) is the most static category with the lowest prestige. Implicitly, being rural reduces the upward mobility opportunity a lot compared to the urban group. According to the literature, people having a low upward mobility prospect should be more in favour of redistribution. However, it is just the reverse case here. This result is comparable to that of the Whyte et al. (2009). This is because the return of physical efforts in farm labour are more straightforwardly perceived than when the comparison is made in the urban group. Rural people tend to believe that poverty is a direct indicator of a lack of effort. Moreover, as argued in Whyte et al. (2009), compared to the urban group, the rural group is a relatively more closed society with much fewer upstart examples so that people living there do not perceive an important level of within-group inequality. The different redistribution schemes applied in rural and urban areas also entails differences in poverty perception. This might be an evidence of perception distortion due to `Hukou` system.

5.6.3 Preference for redistribution

Now let's look at the first column of Table 5.9 which corresponds to the estimate of the preference for redistribution equation. The region dummy effects are not significant in this equation. The birth cohort effects are significant and positive. This means that people who were born later are more supportive of redistribution (Chinese society is changing). Females are more supportive of redistribution (the positive effect of gender upon preference could be emphasized via an indirect channel through *poor-lazy* equation since the two variables are negatively correlated). This effect has also been found in the literature for many different countries. The party membership has no significant direct effect in the preference equation (but a significant negative indirect effect via *poor-misgov* channel) while having a religion belief would decrease the preference for redistribution. Living in couple has no significant effect. Being rural reduces the support for redistribution but this effect is significant only at the 10% level. However, note that we found a very significant effect of the "rural" variable in the other two equations. The "rural" variable plays a very important role in poverty perceptions and thus influences preference for redistribution. Although, the effect of being "rural" is still unclear. To which extent would rural individuals be less in favor of redistribution? Is that due to the segmentation entailed by the rural-urban barrier? We shall discuss this in the next subsection. The number of years of education has no significant effect in the preference for redistribution equation. Clark and d'Angelo (2008a) found that more educated people are less in favour of redistribution (using the BHPS). However, the effect of education upon the preference for redistribution (after controlling for income) can be ambiguous as this has been pointed out in Alesina and Giuliano (2009). Higher educated people could be more altruistic while they could also take into account the potential loss of the education premium entailed by redistributive policies. Although we find no direct effect of education upon preference for redistribution, the significant indirect effect of education (through the *miss-gov* channel) cannot not be neglected.

The income effect is monotone and negative as discussed for the *poor-misgov* equation. The occupation prestige measured by the scaling metric ϕ_k has no significant effect while the comparisons made between the scaling metric of different generations are very significant. This means that people who experienced an upward mobility compared to their parents are less in favour of redistribution. This result is then coherent with that of Clark (2003). Remember that the sign effect for upward mobility between son and father had to be negative in the *poor-lazy* equation while it is again negative in the preference equation and the correlation between these two equations are found to be negative. This could be due to two reasons: i) the correlation parameter is much smaller than the one between *preference for redistribution* and *poor-misgovernment* equations and ii) one who has experienced an upward mobility shall not agree with the idea that poverty is caused by laziness because it would be equivalent to say that his father was lazy.

The last variable *better finance* which captures the subjective measure of the future household financial situation has no significant direct effect. It only enters with a significant negative impact via the *poor-lazy* channel.

5.6.4 Rural-urban segmentation

If the rural-urban barrier which prevents rural areas migrants from reaching the urban areas does exist, then society should be divided into two isolated parts. Thus the barrier should influence in a diverging way preference for redistribution as well as poverty perception of of the two sub-populations. Nevertheless, we cannot ignore the fact that some individuals, while being registered as rural are in fact migrant workers, living and working in urban areas. Most of them are occupying low paid physical jobs in urban areas and they receive much less social benefits compared to the urban residents, see for instance Wong et al. (2007). On the other hand, what they have experienced and seen in urban areas is totally different from what they have seen in their hometown (discrimination and between group inequality). Thus their preferences and perceptions are drifted compared to those who have stayed in rural areas. The change in attitude of individuals who have a dual identity (rural identification and migrant worker status) should modify the interpretation we have of the rural-urban dichotomy, because the group identified as being rural is heterogenous. On the other hand, although the change in status is difficult, we still observe rural-urban identity changes in the data set. How do rural people change their attitudes once they manage to change their rural status for that of urban residents? Would their attitudes converge to those of the original urban residents or the new urban residents simply behave like the migrant workers?

In order to further understand the effect of the rural-urban segmentation, we insert firstly the the dummy variable "being a migrant worker" which corresponds to those who are working in an urban area while still being registered as rural in the year 2006. Table 5.10 reports the estimates of our updated specification with our new dummy variable (we only report the rural and migrant worker variables while other variables remain statistically unchanged. The log-Likelihood value of the new specification has increased from -17 094 to -17 077). We see that the effect of the rural variables are close to the estimates in Table 5.9. However, being a migrant worker would neutralize the positive effect of being "rural" in the *poor-lazy* equation. Thus to say, migrant workers do not hold a higher tendency that poverty is caused by laziness. Meanwhile, migrant workers are much more in favour of redistribution (than those who are simply "rural" or "urban") even if being "rural" reduces the willingness to redistribute compared to urban individuals. The second dummy variable "new urban" captures residents who have recently altered their rural status to the urban one (within the last 10 years). Compared to the other urban individuals (being urban more than 10 years), we found no significant differences although the differences keep the same sign as for the rural group. Eventually, the effect of the rural-urban barrier upon attitudes is strong. The mechanism of differences in attitudes are transmitted through the rural-urban social economic and political differences.

5.7 Conclusion and discussion

In this paper we have discussed the determinants of preference for redistribution along with the subjective perception of the origins of poverty. It

Table 5.10 – attitudes of migrant workers and new urban residents

	Redis. Pref.	Poor-misgov	Poor-lazy
Rural	-0.097^{*} (0.041)	-0.090^{*} (0.040)	0.140^{***} (0.039)
migrant worker	0.175^{**} (0.065)	-0.023 (0.075)	-0.128^{*} (0.059)
New urban	-0.089 (0.081)	-0.070 (0.080)	0.020 (0.069)
N		5417	
Loglik		-17 064	
R		25	

p-value codes: "***" for 0.001, "**" for 0.01, "*" for 0.05 and "." for 0.1.

is obvious that the self-interest variables are not the sole factors at action here, there are also some unobserved factors such as ideology and psychological traits. The preference for redistribution and the poverty perceptions are correlated topics and the causalities are complex. If they are correlated in unobserved ways (mutually endogenous), the standard independent estimation may leads to an inefficiency problem and less information could be provided. One possible econometric model that is designed to capture the conditional correlations of correlated ordinal variables was given in this paper. The performance of this model was proved to be quite reliable and the model is fast to be estimated and easy to apply.

Several evidence have been found in this paper, using the proposed model. First of all, the correlations among preference for redistribution and the poverty perceptions are important. These results provide a proof of the existence of a `sense of justice`. Meanwhile, laziness and misgovernment are not two negatively correlated causes for poverty, at least in the perception of the Chinese people. The correlations also allow us to investigate the direct and indirect effects of explanatory variables in this simultaneous system, e.g. the effect of party membership and being rural have no direct effect upon preference for redistribution but they could have some influences through indirect channels of the poverty perceptions. From the estimates of *poor-misgov* and *poor-lazy* equation, we see that circumstances drift perceptions too. The differences in perceptions between rural and urban group is mainly due to the rural-urban policy barrier. Our results also provide proofs of the most discussed hypothesis, i.e. the "POUM" effect.

In this paper we also discussed the occupation mobility and prestige in China. Our evidence suggests that the widely used EGP occupation category might not reflect the correct social ladder in China, thus in order to use it, one shall be cautious and some readjustments are necessary.

In this paper, we consider only the correlation in errors. People might also want to insert each endogeneous ordinal dependent variable into the other equations, which refers to an endogenous switching system. But this could lead to a serious incoherency problem so that the model requires to be re-identified. This problem has been discussed in Hajivassiliou et al. (2011) for a two equations system, while the solution for cases that have higher dimensions are not clear. Our next step is to extend the model to endogenous switching model and to solve the identification problem.

Chapter 6

Simulation estimation of two-tiered dynamic panel Tobit models with an application to the labor supply of married women: A comment

6.1 Summary

We find that the empirical results reported in Chang (2011b) are contingent on the specification of the model. The use of Heckman's initial conditions combined with observed and not latent lagged dependant variables leads to a counter-intuitive estimation of the true state dependance. The use of Wooldridge's initial conditions together with the observed lagged dependant variable and a proper modeling of censoring provides a much more natural estimate of the true state dependance together with a clearer interpretation of the decision to participate to the labor market in the two-tiered model. We have to stress the usual fragility of estimation of dynamic parameters in panel data models.

0. This paper was co-authored with Michel Lubrano. A short version of this paper has been conditionally accepted by the Journal of Applied econometrics.

6.2 Introduction

Chang (2011b) (Simulation estimation of two-tiered dynamic panel Tobit models with an application to the labor supply of married women, *Journal of Applied Econometrics* 2011) proposed a computationally practical simulation estimator for the two-tiered dynamic panel Tobit model originally developed for cross section data by Cragg (1971). Chang's main contribution is the extension to the panel case with correlated random effects and dynamics. Estimation is undertaken using the GHK (Geweke-Hajivassiliou-Kean) simulator. The one-tiered dynamic Tobit model for panel data and autocorrelated errors is first used for modelling the rich dynamic structure of the labor force participation decision of married women and second for modelling the number of working hours. This initial model is written as:

$$
\begin{aligned}
y_{it}^* &= y_{i,t-1}\lambda + x_{it}\beta + \bar{x}_{i.}\omega + \epsilon_{it} && (6.1) \\
y_{it} &= \max\{y_{it}^*, 0\} && (6.2)
\end{aligned}
$$

with the following error structure

$$
\begin{aligned}
\epsilon_{it} &= d_i + \nu_{it} && (6.3) \\
\nu_{it} &= \zeta\nu_{i,t-1} + u_{it} && (6.4)
\end{aligned}
$$

The two-tiered structure implies that first the probability of participating ($\text{Prob}(y_{it}^* > 0)$) is computed with a first set of parameters (λ_1, β_1, ω_1) while the number of hours worked (i.e. the conditional expectation of y_{it}), conditioned on the decision of participating, is explained by a second set of parameters (λ_2, β_2, ω_2). The other parameters (error variances) are common to the two decisions. See the paper of Chang (2011b) for the interpretation of the coefficients.

Unfortunately, we had a lot of difficulties in reproducing the estimates reported in Chang (2011b). In particular it is strange that the true state dependance parameter λ has a negative value (leading to oscillations) and thus an unexpected sign. Following Heckman (1981b), individuals having experienced an event in the past are more likely to experience that same event in the future. True state dependance has to be disentangled from spurious state dependance which corresponds solely to the individual propensity to capture this effect and which is measured by ζ, the residual autocorrelation parameter. The purpose of this note is to re-estimate Chang's model with different likelihood specifications (different initial conditions) in order to recover a more satisfactory measure of true state dependance.

6.3 Likelihood function and initial conditions

It is important to detail the likelihood function so as to make clear where the point can be. We have first to define the indicator variable I_{it}:

$$I_{it} = 1 \text{ if } y^*_{it} > 0, \text{ and zero otherwise}$$

The likelihood function with fixed initial conditions for individual i is:

$$L_i = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}[f^{(r)}(y_{it}|y_{i,t-1},d_i,x_{it})]^{I_{it}} \times [P^{(r)}(I_{it}=0|y_{i,t-1},d_i,x_{it})]^{1-I_{it}}$$

The latent variables are simulated according to

$$y^{*(r)}_{it} = x'_{it}\beta + \lambda y_{i,t-1} + \bar{x}'_{i.}\omega + A_{t.}(\psi)\eta^{(r)}_{i.}(\psi)$$

where ψ represent the set of all the parameters. The GHK simulator is used to simulate the R replications of η_{it} recursively. The initial conditions can be modeled in different ways. The most simple solution is to suppose that they are fixed and observed leading simply to maximize the above likelihood function. However, this will tend to overstate the true degree of state dependence λ at the expense of the autocorrelation coefficient ζ as noted for instance in Stewart (2006) for the dynamic probit model. The other solution is to suppose that the initial conditions are random and correlated with the individual effects d_i. Heckman (1981a) proposed to model the random initial conditions using an approximation to the reduced form of the model which can be written:

$$y^*_{i0} = z'_{i0}\pi + \theta\bar{x}'_{i.}\omega + \theta d_i + u_{i0}$$

where z_{i0} includes all the exogenous variables plus at least one instrumental variable (see e.g. the implementation in Stewart 2006). Then we can write and maximize the completed likelihood function:

$$\begin{aligned} L_i &= \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}[f^{(r)}(y_{it}|y_{i,t-1},d_i,x_{it})]^{I_{it}} \times [P^{(r)}(I_{it}=0|y_{i,t-1},d_i,x_{it})]^{1-I_{it}} \\ &\quad \times[f^{(r)}(y_{i0}|d_i,z_{i0})]^{I_{i0}} \times [P^{(r)}(I_{i0}=0|d_i,z_{i0})]^{1-I_{i0}} \end{aligned}$$

Chang (2011b) reports using Heckman's approach for the initial conditions, but without detailing their precise specification. In particular, he does not mention either θ or which instrumental variable he has used. Moreover, when using Heckman's initial conditions, it should be more natural to use the lagged latent variable and not the observed one because the reduced form concerns the latent. But this might be at the expense of making more problematic the convergence of the algorithm.

Wooldridge (2005) has proposed a much simpler solution to the treatment of initial conditions in dynamic Tobit models that is much more parsimonious in term of extra parameters and which needs no instrumental variables. Instead of completing the conditional likelihood function by specifying $f(y_{i0}|c_i, \bar{x}_i)$, Wooldridge (2005) proposes to specify $f(c_i|y_{i0}, \bar{x}_i)$ for completing the conditional likelihood function. The method is simple for Probit models as the lagged binary outcomes are of the same nature. For the Tobit model, things are different as underlined by Wooldridge (2005, exemple 2, pages 41-42) as the lagged outcomes have a different scaling depending on censoring. To cope with that, Wooldridge introduces a $\mathbf{g}(y_{t-1})$ function which gives the following equivalent specification for our model:

$$c_i = \bar{x}_i'\omega + d_i + y_{i0}\delta_1 \mathbb{I}(y_{i0} > 0) + \delta_2 \mathbb{I}(y_{i0} = 0)$$

And of course as the lagged endogenous variable is treated as observed, the formulation of Wooldridge (2005, page 49) leads to adopting two different λ's when censored or not censored.

$$y_{it}^* = x_{it}'\beta + y_{i,t-1}\lambda_1 \mathbb{I}(y_{i,t-1} > 0) + \lambda_2 \mathbb{I}(y_{i,t-1} = 0) + c_i + u_{it}$$

Wooldridge (2005) did not consider the two-tiered model, so we have to detail the extension that we used for the c_i. For the participation equation, we have:

$$c_{1i} = \bar{x}_i'\omega_1 + d_i + y_{i0}\delta_{11} \mathbb{I}(y_{i0} > 0) + \delta_{12} \mathbb{I}(y_{i0} = 0)$$

while for working hours equation

$$c_{2i} = \bar{x}_i'\omega_2 + d_i + y_{i0}\delta_{21} \mathbb{I}(y_{i0} > 0) + \delta_{22} \mathbb{I}(y_{i0} = 0)$$

Of course, we can test the restriction $\delta_{11} = \delta_{21}$ and $\delta_{12} = \delta_{22}$. Comment on ζ which captures spurious state dependance (the individual propensity to capture true state dependance). With different δs, we allow for a better modeling on the way individuals capture state dependance.

6.4 Checking the empirical results

We have re-programmed (with the free software R) the one-tiered and two-tiered dynamic panel Tobit models using the indications in Chang (2011b) completed by Chang (2011a) and Stewart (2006). We first focus on the one-tiered version which corresponds to the second column of Table III, page 866 of Chang (2011b). Following the author's indications, we used 10 draws of the uniform random variable ξ_{it} which is the basic ingredient in the GHK simulator used to dynamically generate the truncated error term of the model. We have to keep the same numbers for ξ_{it} in order to make results comparable between the different models. We have tried three different options for the initial conditions. First, we used fixed initial conditions as the most immediate solution. Then, we implemented Heckman's random initial conditions based on a reduced form. That means using an instrumental variable for the initial state. We used the years of education of the husband. We finally implemented Wooldridge's initial conditions with censoring for the lags.

6.4.1 Optimization strategy

We have found that convergence of the algorithm clearly depends on how parametric constraints are introduced and how the optimization is proceeded. Chang used the optimization routine `constrOptim` which introduces linear inequality constraints using an adaptive barrier algorithm and then passes arguments to a `BFGS` algorithm. The use of the `BFGS` algorithm requires the specification of an analytical gradient function in that case (otherwise, the `Nelder-Mead` algorithm has to be used). The package `optim` of `R` implements the `L-BFGS-B` algorithm of Byrd et al. (1995), allows box constraints and does not require the specification of an analytical gradient function. An analytical gradient function is much more efficient, but can give rise to movements in a wrong direction in case of programming errors. Note that the performance of routine `optim` is not very reliable for MLE, same as `constrOptim`, which gives the same output. The main reason of that is because the routine `optim` (or `constrOptim`) does not calculate the numerical gradients in a proper way. These issues may lead the calculations towards wrong directions. Instead, the routine `maxLik` seems to be a better choice. In fact, by providing an symmetric numerical gradient formula would allow `optim` to obtain the same result as `maxLik`.

Another issue concerns the choice of the maximization method. The standard errors are calculated via the hessian matrix. Note that the standard calculation procedure of the hessian might fail (unprecise or singular) if the shape of the log-likelihood function is flat around the maximum. We are in this case and we shall show it in the following discussions. One alternative hessian approximation would refer to the `BHHH` (Berndt et al. 1974) based on the information equality approach. Using `BHHH` the hessian matrix is guaranteed to be positive semi-definite and thus the estimation of the standard errors is not sensitive to the shape of the log-likelihood function. As it has been argued in Doan (2010, section 9.4), `BHHH` could be used only for a final iteration to get the estimation precision once `BFGS` is converged in case that the estimated hessian matrix is singular.

Which parameters have to be constrained? We should note that if σ_d and σ_u have to be positive, this positivity constraint is directly fulfilled as σ_d and σ_u appear only as squares in the likelihood function. Otherwise, only the stationarity assumption $|\zeta| < 1$ appears to be crucial for the success of the optimization process.

6.4.2 Empirical results

We reproduce in column one of Table 6.1 the results of Chang's one-tiered Correlated RE+AR(1) model. Using those starting values completed by 1 for θ, 0 for the instrument parameter and (600, 800) for the two error components, we reached the maximum of the likelihood function using the Heckman's initial conditions in 150 iterations. On average, parameter values are very close, within one standard deviation most of the time. The standard errors are generally slightly smaller than those of Chang. But the state dependance parameter is negative. If we constrain $\theta = 1$ and use no instrumental variable (as was possibly done in Chang's paper), the likelihood function is lower with -91 771 and the algorithm converges in 98 iterations. If we use fixed initial conditions, the algorithm converges in 106 iterations, with slightly lower standard deviations and now a positive and significant true state dependence parameter, a slightly lower value for ζ and a better likelihood value than the one obtained when restricting θ to 1 and excluding the instrument in the Heckman's specification.

Let us now implement the Wooldridge's specification for the initial conditions, allowing in a very simple way for a different parameter λ (δ) when the lagged (initial) dependent variable is censored or not as recommended in Wooldridge (2005). In the fourth column (estimated via `Optim` with `BFGS`), the log-likelihood value is now much higher, and the dynamics has changed a lot. There is a much more important positive true state dependance measured by λ_1 for uncensored past working hours while the value of ζ becomes negative and insignificant in this column. The intra-class correlation has also changed with a much lower value. For censored lagged endogenous, the coefficient is negative, significant with a scaled value of -0.262 (scaled for the mean of positive values of hours of work in 1984). In the last column we give the final estimation via `maxLik` with `BFGS` + `BHHH` and the same model specification. The standard errors in the last column are almost always smaller than those reported in the fourth column. Coefficient for ζ has even changed the sign and appears to be significant. And we shall notice that the log-likelihood value is also improved.[1]

What are our main conclusions? First, we could not reproduce Chang (2011b) results in a satisfactory way as we had no clear clue on how to empirical implement Heckman's initial conditions with Chang data set. Second, the choices made for the specification of the initial conditions are of prime importance for measuring accurately the true state dependance. The Heckman's approach should ideally be combined with a latent lagged endogenous variable. Wooldridge's specification is much simpler as it allows naturally for an observed lagged endogenous variable, even if a distinction has to be made when that variable is censored or not. It seems to be the most satisfactory approach. The measure it provides for true state dependance is in accordance with intuition and it reduces the importance of individual random effects, the latter having a much lower variance. Third, the convergence of `BFGS` and `BHHH` are quite similar (from $-91\ 111$ to $-91\ 108$) while the latter one provides better estimation precision. The choice of the optimization routine matters. We found that with any specification, `Optim` with `BFGS` does not lead us to the optimum. For instance, in the second and the third column, we estimated two specifications that are close to the specification of Chang and the iteration ends up within reasonable terms (starting values are drawn from the result of Chang). If we change the specification, it takes 1 605 iterations to reach the "maximum". This would be a clear indication that our log-likelihood function has a flat shape and the performance of `Optim` is not reliable.

We obtained similar results with the two-tiered model.[2] Since we have one set of parameters for each equation and we distinguish censored and uncensored events, we have four parameters for λ and δ. We obtained positive and significant effects for the uncensored λ_1, and a negative and significant (positive and insignificant) λ_2 for the censored values in the participation (working hours) equation. This would mean that the more hours individual worked (at least participated) in $t-1$, the more likely that this individual would continue to work in t with more working hours compared to others. However, individuals who report zero hour of work (censored) in $t-1$ are less likely to work in the next period while this event has no significant effect upon working hours decision once that individual find a job. Meanwhile,

1. We have tested all specifications via `maxLik` with `BFGS` or `BFGS` + `BHHH` and they work always better than `Optim` or `constrOptim` with `BFGS`. In this note we only report the estimation using same optimization strategy as Chang (we found no difference between `Optim` and `constrOptim` in our case) except for our final one in order to do the comparison.

2. We only report the estimation result obtained via `maxLik` with `BFGS` + `BHHH`. Similar to the one-tiered case, the estimation via `Optim` with `BFGS` ends up with lower log-likelihood value and larger standard errors. It takes more than six hours to "converge" with almost 900 iterations.

Women who start without a job consistently tend to have no job, vice versa for women who start with a job. The same for working hour equation, except that staring with zero hour (censored) now has a positive and significant effect on subsequential working hour decision. Similarly to the ζ reported in the final estimation of the one-tiered model, it has a significant negative effect which refers to some oscillations in unobservable ways. The log-likelihood function has a much higher value. Clearly this model is an improvement over the one-tiered model not only because that the log-likelihood value is higher but also it relaxes many constraints that allows to capture the asymmetric effects between two equations. The Wooldridge's specification provides a better fit and a likelihood which is easier to evaluate than that of the case with fixed initial conditions for the two-tiered model (unreported results). Some coefficients are poorly identified: C35 in the first tier, education and C613 in the second tier. However, with our specification, the number of children has a clear impact on the decision to participate. Mothers with young babies do not participate, with mid-age kids the effect is not significant while with kids between 6 and 13 mothers are more likely to participate. This story is more intuitive than the story illustrated by Chang's results.

Starting values are quite difficult to choose in any optimization problem. We choose to start from the values given by Chang, using trials and errors for the other parameters, the aim of the game being to get a feasible value for the likelihood function before launching the optimization. This is not too complex for scale free parameters. For the standard deviations σ_u and σ_d, we can base the starting values on the empirical standard deviation of the endogenous variable. The starting values of the two-tiered can be based on the results of the one-tiered model. We could reduce the number of iterations with carefully chosen starting values, but the final result was not different. In fact choosing starting values is not so much different as eliciting a prior in a Bayesian framework (see e.g. Bauwens et al. 1999, Chap. 4). One important question is also the choice of R, the number of replications for the GHK simulator. Stewart (2006) reports the dynamic Probit model experiments and suggests $R \geq 100$. We have used `Optim` with `BFGS` and $R = 10$ leading to an execution time of 4 hours for the one-tiered model. In our final estimations of one-tiered and two-tiered model, due to the optimization strategy we have chosen, the calculations are much better and faster (both within 15 minutes). We found the same results between $R = 10$ and $R = 100$.

6.5 Conclusion

A dynamic Tobit model was chosen to model female labor supply decisions, the decision to work or not and the decision of the number of hours to work. There is clear censoring in the data and we found that modeling the censoring should be applied for both the dependent variable and its lagged value. When Heckman's solution is applied for the initial conditions, it might seem natural to include the lagged latent variable in the model at the expense of having perhaps a less tractable likelihood function with real data. If the lagged observed is included as it was in Chang (2011b), this is equivalent to ignoring the censored values because $0 \times \lambda$ mixes with $y_{i,t-1} \times \lambda$ and leads to a distortion of the dynamics. We believe that a negative coefficient for both censored and uncensored lagged dependent variable would lead to a conclusion of oscillations in the true state dependence which is not very intuitive. Wooldridge (2005) clearly states that in the case of a Tobit model his proposal should lead to two dynamic coefficients, one for the censored lag, one for the observed lag. This is what we did in this note. Using Wooldridge's specification reduces a lot the intra-class correlation and thus leads to a less important value for ζ compared to Chang's reported results. With Heckman's, we have to model y_0 using a not so good approximation, while with Wooldridge's, we use y_0 to model the individual effects, thus reducing greatly the variance σ_d^2 and the autocorrelation of the ϵ_{it}. Moreover, by relaxing the symmetric constraints of censored/uncensored within two tiered framework, we manage to find richer dynamic information which is not allowed by the one-tiered model. With the two-tiered model, Wooldridge's approach provides a better estimate of the effect of the number of children between 6-13 on the decision to participate, which makes the model more convincing.

Table 6.1 – Correlated RE+AR(1) one-tiered dynamic panel Tobit model of married women's labor supply

	Chang (2011)	Heckman initial conditions	Fixed initial conditions	Wooldridge with censoring	
Routine	ConstrOptim	Optim	Optim	Optim	maxLik
max. mtd.	BFGS	BFGS	BFGS	BFGS	BFGS+BHHH
Const	−83.49	−83.73	−84.25	−94.95	3.21
	(321.7)	(317.05)	(261.3)	(156.50)	(133.17)
Edu	79.48	76.95	70.87	29.43	14.66
	(12.98)	(12.19)	(11.70)	(5.82)	(4.70)
Race	131.6	131.36	130.8	108.30	63.01
	(57.23)	(57.34)	(55.73)	(27.06)	(22.62)
Age	50.34	46.52	37.06	18.48	23.01
	(15.19)	(14.04)	(11.24)	(7.10)	(5.97)
Age^2	−0.803	−0.78	−0.650	−0.35	−0.37
	(0.188)	(0.17)	(0.141)	(0.08)	(0.07)
Hinc	−1.509	−1.64	−1.622	−1.49	−1.41
	(0.435)	(0.44)	(0.438)	(0.39)	(0.47)
C12	−140.3	−141.00	−141.8	−146.07	−64.63
	(24.12)	(19.80)	(19.13)	(17.33)	(14.27)
C35	−114.6	−114.78	−115.3	−61.79	−30.58
	(21.07)	(19.60)	(18.22)	(15.19)	(13.89)
C613	−26.15	−26.22	−26.20	6.64	17.87
	(16.03)	(15.72)	(14.49)	(11.04)	(10.48)
$\overline{Hinc}$	−7.010	−6.22	−5.696	−1.30	−1.14
	(2.002)	(1.15)	(1.092)	(0.59)	(0.59)
$\overline{C12}$	−264.1	−264.24	−264.4	−249.40	−279.81
	(146.3)	(170.46)	(152.6)	(73.11)	(61.09)
$\overline{C35}$	−290.8	−290.97	−291.2	−240.21	−9.64
	(143.6)	(167.16)	(153.9)	(72.89)	(59.78)
$\overline{C613}$	−81.74	−81.95	−82.52	47.14	9.40
	(43.05)	(48.67)	(44.86)	(23.31)	(19.99)
λ	−0.054	−0.096	0.108		
	(0.019)	(0.013)	(0.040)		
λ_1				0.44	0.52
				(0.02)	(0.01)
λ_2				−378.70	−500.91
				(23.47)	(16.93)
σ_u		597.77	588.4	573.47	583.97
		(4.51)	(4.368)	(4.16)	(2.26)
σ_d		799.87	817.8	367.81	277.80
		(25.78)	(35.83)	(16.72)	(9.87)
ζ	0.613	0.70	0.515	0.03	−0.05
		(0.014)	(0.042)	(0.02)	(0.01)
θ		0.79			
		(0.027)			
heduc		−13.08			
		(2.00)			
δ_1				0.25	0.16
				(0.02)	(0.01)
δ_2				−113.54	−180.61
				(35.92)	(26.33)
ρ	0.721		0.659	0.290	0.184
Log-Like	-92 091	-91 734	-91 764	-91 171	−91 111 $(BFGS)$
Iterations		150	106	1 605	145 $(BFGS)$
Log-Like					−91 108 $(BHHH)$
Iterations					42 $(BHHH)$

Standard errors are shown in parentheses. The first column corresponds to the published results of Chang. The second column implements the model of Heckman with an approximate reduced form using the husband education level as an instrumental variable. It should correspond to the replication of Chang. The third column gives the estimation with fixed initial conditions. The last two columns correspond to the full implementation of Wooldridge (2005) estimated via different optimization strategy.

Table 6.2 – Correlated RE+AR(1) two-tiered dynamic panel Tobit model of married women's labor supply

	Chang (2011)		Replication	
Routine	ConstrOptim		MaxLik	
max. mtd.	BFGS		BFGS+BHHH	
	Particip.	Hours	Particip.	Hours
Const	−572.6 (302.4)	378.1 (310.0)	−194.45 (150.34)	444.29 (102.39)
Edu	77.28 (11.27)	30.94 (13.22)	28.15 (4.58)	−3.37 (3.09)
Race	73.25 (49.94)	239.7 (48.30)	−20.40 (22.71)	78.62 (17.04)
Age	59.50 (14.85)	66.37 (14.74)	19.08 (7.18)	12.74 (4.80)
Age^2	−0.852 (0.181)	−0.928 (0.185)	−0.30 (0.08)	−0.19 (0.05)
Hinc	−1.682 (0.437)	−1.100 (0.509)	−1.17 (0.72)	−0.97 (0.46)
C12	−92.75 (32.50)	−121.0 (24.06)	−63.69 (23.58)	−22.50 (15.14)
C35	−66.68 (28.89)	−102.1 (21.52)	6.25 (22.29)	−19.78 (12.48)
C613	12.83 (23.34)	−31.51 (15.60)	50.82 (17.66)	10.97 (9.91)
$\overline{Hinc}$	−6.164 (1.532)	−7.958 (2.205)	−0.48 (0.79)	−1.21 (0.57)
$\overline{C12}$	−309.7 (134.5)	−121.9 (139.1)	−153.65 (61.96)	199.96 (45.51)
$\overline{C35}$	−207.4 (131.8)	−356.9 (143.5)	−11.49 (58.11)	−13.92 (42.72)
$\overline{C613}$	−74.40 (42.16)	−86.74 (40.37)	−0.42 (25.35)	−14.72 (15.37)
λ	−0.102 (0.042)	−0.052 (0.025)		
λ_1			0.38 (0.02)	0.53 (0.01)
λ_2			−448.87 (20.62)	28.33 (19.20)
σ_u			492.23 (1.96)	
σ_d			173.84 (6.83)	
ζ	0.608		−0.09 (0.01)	
δ_1			0.03 (0.01)	0.12 (0.01)
δ_2			−198.21 (24.50)	54.96 (18.49)
ρ	0.574		0.110	
Log-Like	-91 331		−90 288 $(BFGS)$	
Iterations			163 $(BFGS)$	
Log-Like			−90 277 $(BHHH)$	
Iterations			52 $(BHHH)$	

Standard errors are shown in parentheses.

ïż£

Chapter 7

General conclusion

In this dissertation we have studied many different economic topics by using subjective data sets and subjective approaches.

In the second chapter we found that individuals perceive different aspects of the reference income distribution, as both the mean level and the dispersion are important. Different people draw different information from the others who are around them, either in a positive and maybe excited way, depending if they see the success “people like me” as a personal possibility future success or in a negative and depressed way if they consider the gap between themselves and “people like me” with envy and anyway as too difficult to overcome.

In the third chapter we redefined the International Poverty Line. The main conclusion is that even in the least developed country group, we found a significant consumption elasticity although it is much smaller than the elasticity found for the more developed group. This is not a surprise that we should not limit the basic human needs to calories but that we have also to consider whether people have access to resources that ensures a decent life experience. The new definition of an international poverty line is that of the mean level of consumption multiplied by the consumption elasticity found in the least developed group. Using this new definition of an international poverty line, we found a posterior density that has a posterior mean which is higher than the classical estimate reported by the World Bank and a posterior standard deviation which is much smaller than its classical counterpart.

In the fourth chapter we explained the preference for redistribution by two types of dynamics, job dynamics and income dynamics. We found that both dynamics have important impacts. Depending on their job status, the set of determinant might change. We found a clear evidence in favour of the ”POUM” effect, but the latter can be greatly weaken by individual risk aversion (probability of losing a job).

In the fifth chapter, we relate the preference for redistribution to two subjective variables that correspond to questions related to the perception of the origin of poverty, more precisely the role of “efforts” and “circumstances”. By using a new econometric model (trivariate ordered probit), we managed to find a clear correlation between preference for redistribution and the two perception variables. The result provides a clear evidence of the role of “efforts” and “circumstances” which are shown not to be mutually exclusive. That was not proved quantitatively before.

It is clear that until now we had only a very limited understanding of what and how people feel and think. The direct information provided in current data sets is also limited, thus in order to explore this information, we need to use and construct different type of models in order to properly measure dynamics, correlations, etc. As dynamic information is extensively used in this dissertation, we would like to point out some limitations of the methodology that we have used in this dissertation. For example in chapter two we used the dynamic multinomial logit model to measure job and income dynamics. One important issue is that the initial status is not fixed as we supposed, but results from a proper random process, possibly stationary. In the last chapter of this dissertation, we showed the importance of correctly specifying the initial conditions in a dynamic Tobit model. So clearly more work is needed to extend our results of chapter four.

In this dissertation we have also tried to consider several subjective variables at a time in a simultaneous model (chapter five). The basic idea was to show i) what are the correlations between these subjective variables and ii) if they are correlated, ignoring some important information would lead to estimation problems, especially for the discrete variables. Because of complicated coherency conditions, we only considered a joint reduced form model. More work is needed in order to be able able to propose a full simultaneous system.

Appendix A

Appendices

A.1 Chapter 2

A.1.1 CASMIN levels

CASMIN classification as given in the BHPS documentation. For more details, see Muller (2000). These nine classes were used to determined reference groups and reference income. Table A.1 gives their definition and frequency in the sample for 2008. Individuals with missing values were deleted.

Table A.1 – CASMIN lelvels, last wave

CASMIN	Education level	Value	Frequency	%
1a	none	1	2532	19.2
1b	elementary	2	503	3.8
1c	basic vocational	3	1131	8.6
2b	middle general	4	2257	17.1
2a	middle vocational	5	664	5.0
2c-gen	high general	6	1186	9.0
2c-voc	high vocational	7	741	5.6
3a	low tertiary	8	2218	16.8
3b	high tertiary	9	1956	14.8

A.1.2 Metropolitan areas

These nineteen areas were used to determine secondary reference groups in order to measure sensitivity to overall inequality. Table A.2 gives their definition and sample frequency for 2008. The last wave has no missing value.

Table A.2 – Metropolitan areas, last wave

Zone	Code	Frequency	%
Inner London	1	117	1.4
Outer London	2	242	3.0
R. of South East	3	881	10.8
South West	4	450	5.5
East Anglia	5	225	2.8
East Midlands	6	401	4.9
West Midlands Conurb	7	145	1.8
R. of West Midlands	8	249	3.1
Greater Manchester	9	172	2.1
Merseyside	10	118	1.4
R. of North West	11	234	2.9
South Yorkshire	12	140	1.7
West Yorkshire	13	158	1.9
R. of Yorks and Humber	14	158	1.9
Tyne and Wear	15	102	1.3
R. of North	16	184	2.3
Wales	17	1427	17.5
Scotland	18	1497	18.4
Northern Ireland	19	1244	15.3

The frequency of missing values is very small in other waves. Assuming that households are not moving frequently, whenever we had a missing value in waves L to Q, we assigned the location declared in the next wave. Note the numerical importance of the last three regions.

A.1.3 Comparing two independent regressions

We want to compare two identical regressions, labeled 1 and 2, which are run on two different samples. For comparing all the coefficients together, we use the following Wald test:

$$(\Theta^1 - \Theta^2)'(\Sigma^1_\Theta + \Sigma^2_\Theta)^{-1}(\Theta^1 - \Theta^2) \sim \chi^2(k) \tag{A.1}$$

where k is the number of estimated coefficients.

For comparing only two individual coefficients, we test that their difference is zero with a $t-$test:

$$z = (\beta_1 - \beta_2)/\sqrt{\sigma_1^2 + \sigma_2^2}.$$

Note a similar approach in Ferrer-i-Carbonell (2005).

In section 2.4.5, we want to compare two ratios of coefficients. We can still use a t-test, but we have to use the Delta method to compute the variance of a ratio. From Cramer (1946, pp. 353-359), we know that the variance of a ratio $h = \beta_1/\beta_0$ can be approximated by:

$$\text{Var } h \simeq (\frac{\partial h}{\partial \beta_1})^2 \text{Var } \beta_1 + 2\frac{\partial h}{\partial \beta_1}\frac{\partial h}{\partial \beta_0}\text{Cov}(\beta_1, \beta_0) + (\frac{\partial h}{\partial \beta_0})^2 \text{Var}\beta_0$$

which reduces to

$$\text{Var } \frac{\beta_1}{\beta_0} \simeq \frac{1}{\beta_0^2}\text{Var } \beta_1 - 2\frac{\beta_1}{\beta_0^3}\text{Cov}(\beta_1, \beta_0) + \frac{\beta_1^2}{\beta_0^4}\text{Var}\beta_0.$$

A.2 Chapter 3

A.2.1 Simulation of a bivariate density using a grid

Let us consider a bivariate posterior density:

$$\pi(\phi, \theta|y) = \pi(\phi|\theta, y) \times \pi(\theta|y)$$

We know the analytical form of the joint density $\pi(\phi, \theta|y)$, but neither its marginal $\pi(\theta|y)$ nor its conditional $\pi(\phi|\theta, y)$. We want to draw random numbers for the joint posterior density. To do so, we are first going to evaluate this bivariate density on a grid, filling a matrix F where the rows will corresponds to θ and the columns to ϕ. From this matrix of points, we can determine numerically the marginal density $\pi(\theta|y)$ by summing over the columns. Using this marginal density and using the inverse transformation method, we can draw a value for θ. For a given draw of θ, we have to find the corresponding conditional density $\pi(\phi|\theta, y)$ as a row of matrix F. Of course, the draw will not correspond exactly to one of the predetermined point of the grid in θ. So we shall have to proceed by linear interpolation between two lines.

1. Compte numerically the cumulative and then use the inverse transformation method to draw θ^j from $\pi(\theta|y)$
2. Find the two nearest points of θ^j on the grid of θ, denoted as θ^{j-} and θ^{j+}
3. Calculate the differences: $a = \theta^j - \theta^{j-}$, $b = \theta^{j+} - \theta^j$ and $c = |\theta^{j+} - \theta^{j-}|$
4. Obtain the conditional posterior densities $\pi(\phi|\theta^{j-}, y)$ and $\pi(\phi|\theta^{j+}, y)$ from the joint posterior matrix F

5. Compare each point of the two above conditional posterior densities in order to get $\pi(\phi|\theta^j, y)$ by line interpolation:

$$\sum_{k=1}^{k} \pi(\phi_k|\theta^j, y) = 1$$

$$\pi(\phi|\theta^j, y) = \begin{cases} \pi(\phi_k|\theta^{j-}, y) + a \times (\pi(\phi_k|\theta^{j+}, y) - \pi(\phi_k|\theta^{j+}, y))/c & \text{if } \pi(\phi_k|\theta^{j+}, y) \geq \pi(\phi_k|\theta^{j-}, y) \\ \pi(\phi_k|\theta^{j+}, y) + b \times (\pi(\phi_k|\theta^{j-}, y) - \pi(\phi_k|\theta^{j+}, y))/c & \text{otherwise} \end{cases}$$

knowing that ϕ_k is the kth point on the grid of ϕ

6. Compte numerically the cumulative and then use the inverse transformation method to draw ϕ^j from $\pi(\phi|\theta^j, y)$

7. Store the jth joint draw : (θ^j, ϕ^j)

A.3 Chapter 4

A.4 Markov processes

We consider a discrete Markov chain with constant transition matrix P with typical element $p_{ij} = \text{Prob}(\text{state} = j \text{ in } t+1| \text{ state} = i \text{ in } t)$, formalizing the transition between K ordered income states $y_1 < y_2 < \cdots < y_K$. The vector of the K states is characterized by a probability vector π_t (marginal distribution) evolving over time according to $\pi_{t+1} = \pi_t P$. P is said to be regular if after a sufficiently large time span, every state is visited, or in other terms that P^t has all its elements which are strictly positive. Under the regularity condition for P, the equilibrium vector $\pi_e = \pi_e P$ exists and is unique. [1] Expectations can be formed with this model meaning that the expected distribution or future income opportunities are given by $\pi_t = \pi_0 P^t$. π_t denotes the proportion of individuals in each of the K states after t periods.

A.4.1 Mobility indices

A transition matrix P has K independent rows. Each row indicates the probability to change from status i to status j the next period and sums to

1. It can be shown, see e.g. Guilbaud (1977) that the equilibrium vector is given by

$$\pi_e = [(P - I)(P - I)' + JJ']^{-1} J$$

where I is the identity matrix and J the vector of ones.

1. Overall mobility can be summarized using a mobility index, that of Prais (1955) being the most simple one:

$$M_P(P) = (k - \mathrm{Tr}(P))/(k-1).$$

$M_P(P) = 0$ is perfect immobility while $M_P(P) = 1$ is perfect mobility. According to this index, $P = I_k$ represents perfect immobility while perfect mobility is more complex to define. M_P represents the average normalized distance to the identity matrix. The usual way of representing perfect mobility is to assume that all the rows are equal and correspond to the stationary distribution associated to P. So that perfect mobility is not the same between two different countries for instance. A more extreme definition of perfect mobility would be to assume that $P = [p_{ij} = 1/k]$.

A.4.2 Monotone transition matrices

A class of transition matrices will be particularly interesting for characterizing income mobility. This is the class of monotone transition matrices extensively detailed in Dardanoni (1993) and Dardanoni (1995) for the economic literature or Conlisk (1990) for the sociological literature. The notion simply says that an individual starting from income status $i+1$ faces a better lottery than an individual starting from status i. Formally, we have the following definition as reported in Benabou and Ok (2001b):

Definition 1. *A transition matrix is said to be monotone if each row $i+1$ stochastically dominates its predecessor i, which means:*

$$\sum_{j=1}^{l} p_{1,j} \geq \sum_{j=1}^{l} p_{2,j} \geq \cdots \geq \sum_{j=1}^{l} p_{K,j} \qquad \forall l = 1, 2, \ldots, K-1. \tag{A.2}$$

When a transition matrix is monotone, this means that an individual with status $i+1$ faces a better expected future income or opportunity than an individual with initial status i. Monotonicity is a convenient property for comparing transition matrices. For instance when $P \geq \tilde{P}$, the dominance relation extends to equilibrium vectors $\pi_e \geq \tilde{\pi}_e$ only when both P and $\tilde{P}$ are monotone.

Dardanoni and Forcina (1998) provide another definition of stochastic ordering of random variables. Expressed in our notations, their condition is

$$\frac{p_{1,j}}{p_{1,j+1}} \geq \frac{p_{2,j}}{p_{2,j+1}} \geq \cdots \geq \frac{p_{K,j}}{p_{K,j+1}} \qquad \forall j = 1, 2, \ldots, K-1. \tag{A.3}$$

They show that (A.3) is a necessary (but not sufficient) condition for (A.2) to hold.

A.4.3 Equality of opportunity

For Benabou and Ok (2001b), mobility is interesting because it may provide equality of opportunity, which means that the opportunity of an individual are independent of his initial state or social origin. A transition matrix will be gauged according to its capacity to equalize opportunities or expected future incomes. This means that the expected income of the poor has to increase more than the expected income of the rich. Mobility is a form of stochastic redistribution.

The distribution of conditional expected income or opportunity for an individual in state i is given by

$$e_P(y_i) = e_i = \sum_{j=1}^{K} p_{ij} y_j.$$

Benabou and Ok (2001b) provide the following definition which is used in Benabou and Ok (2001c):

Definition 2. *A transition matrix P is more equalizing than a transition matrix Q if the conditional distribution of expected incomes generated by P is more equal than that generated by Q for all initial distributions π. We note $P \succcurlyeq_{eq}^{y} Q$.*

Definition 3. *A monotone transition matrix P is progressive if it is more equalizing than the identity matrix. We note $P \succcurlyeq_{eq}^{y} I_K$, the identity matrix I_K formalizing immobility.*

Benabou and Ok (2001b) give a theorem which allows to characterize both progressivity and compare transition matrices in term of opportunity equalization.

Theorem 2. *Let us consider two monotone transition matrices P and Q between n income states $y_1 < y_2 < \cdots < y_K$. The monotone transition matrix P is more equalizing than the monotone transition matrix Q, a statement noted $P \succcurlyeq_{eq}^{y} Q$, if we have:*

$$\frac{e_P(y_1)}{e_Q(y_1)} \geq \frac{e_P(y_2)}{e_Q(y_2)} \geq \cdots \geq \frac{e_P(y_K)}{e_Q(y_K)}.$$

The condition for progressivity is obtained by setting $Q = I_K$, so that to obtain:

$$\frac{1}{y_i} \sum_{j=1}^{K} p_{ij} y_j \geq \frac{1}{y_{i+1}} \sum_{j=1}^{K} p_{i+1,j} y_j, \qquad i = 1, \ldots, K-1. \tag{A.4}$$

Benabou and Ok (2001b) study the mobility of a process as a function of (P, y), whatever the value of π, the marginal distribution of the initial states or the equilibrium distribution. But their analysis is always conditional on the definition of the states, the y, arguing that income mobility should characterize both P and y and not just simply P. So if we want to compare two transition matrices P and Q, they have to be both size transition matrices where the boundary classes y are determined exogenously.

On the contrary, Dardanoni (1993) ranks two transition matrices for a given equilibrium distribution π, whatever the value of the y. This means that P is preferred to $\tilde{P}$ if:

$$\sum_{i,j}^{k,l} \pi_j p_{ij} \leq \sum_{i,j}^{k,l} \pi_j \tilde{p}_{ij} \qquad \forall k, l = 1, \cdots, K-1.$$

For quantile transition matrices, of course all the $\pi_j = 1/k$.

A.4.4 Testing for homogeneity

A Markov transition matrix is usually estimated by maximum likelihood which is shown to correspond to (see the seminal paper of Anderson and Goodman 1957 or e.g. the statistical appendix in Boudon 1973):

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{\sum_j n_{ij}(t)} \tag{A.5}$$

where $n_{ij}(t)$ the number of individuals in state i at time $t-1$ moving to state j at time t. When there are more than two periods and if the process is homogenous, the maximum likelihood estimator is obtained by averaging the $\hat{p}_{ij}^t$ obtained between two consecutive periods. This is the procedure which was followed to estimate P in Table 4.3. Of course, this estimator is not at ease when the panel is incomplete (see e.g. Sherlaw-Johnson et al. 1995 for an example using the EM algorithm to cope with missing observations).

A test for homogeneity is based on a χ^2 test as first derived in Anderson and Goodman (1957). It corresponds to the following hypothesis

$$\begin{array}{lcll} H_0 & : & p_{ij}(t) = p_{jk} & \forall t \\ H_1 & : & p_{ij}(t) \neq p_{ij} & \exists t \end{array}$$

Then a Likelihood ratio test is:

$$Q = 2 \sum_{t=1}^{T} \sum_{i=1}^{K} \sum_{j=1}^{K} n_{ij}(t) \log \frac{\hat{p}_{ij}(t)}{\hat{p}_{ij}} \sim \chi^2 \left((T-1)K(K-1)\right),$$

assuming that all the $p_{ij}(t)$ are strictly positive. When one $n_{ij}(t)$ is equal to zero, the corresponding values are just not included in the test.

H_0 can be tested against H_1 using a dynamic multinomial logit model with time dummies. Homogeneity corresponds to the case where the time dummies are not needed.

A.4.5 Testing for regularity or monotonicity

Monotonicity mean that future income is an increasing function of current income. Progressivity means that the income of the poor is increasing more quickly than the income of the rich so that mobility is equivalent to stochastic redistribution. So monotonicity is usually a prerequisite condition for progressivity. Testing for monotonicity means that we have to verify a set of $(K-1)^2$ inequalities defined in (A.2). Testing for progressivity means that we have to test for a set of $K-1$ inequalities defined in (A.4).

Regularity means that it is better to start from state $i+1$ than from state i. The test reported in Dardanoni and Forcina (1998) are quite complex, except that corresponding to test the null H_0 : *the stochastic ordering does not hold* against the alternative H_1 *the stochastic ordering holds strictly.* See their section 4.6 for more precisions. There is a simple way to test the necessary condition (A.3). We first have to note that this condition is equivalent to

$$\frac{p_{ij}}{p_{i+1,j}} \geq \frac{p_{i,j+1}}{p_{i+1,j+1}}.$$

Then, in the dynamic multinomial logit model (4.5), we have that $\exp(\gamma_{jk})$ receives a special interpretation as given in (4.10). So that finally, the above probability ratio can be expressed as a ratio of the $\gamma's$ such that the necessary ranking condition is equivalent to

$$\frac{\exp(\gamma_{ij})}{\exp(\gamma_{i+1,j})} \geq \frac{\exp(\gamma_{i,j+1})}{\exp(\gamma_{i+1,j+1})}$$

or again as the exponential is a monotone increasing function

$$\gamma_{ij} + \gamma_{i+1,j+1} - \gamma_{i,j+1} - \gamma_{i+1,j} \geq 0, \qquad \forall i, j = 1, \ldots, K-1.$$

Let us now define a vector d of dimension $(K-1)^2$ obtained by letting i, j vary from 1 to $K-1$. Monotonicity means that $d \geq 0$, means means testing simultaneously for $(K-1)^2$ inequalities. This type of simultaneous test has been investigated in the literature (see e.g. Dardanoni and Forcina 1998, but also Davidson and Duclos 2000 in a different context). The various test reported there are quite complicated, first for describing the constrained parameter space and second for their asymptotic distribution. However, as anyway (A.3) is just a necessary condition, it is sufficient to find one occurrence where the inequality is violated. So if ω_j is the standard deviation of d_j, we could decide to test for

$$\min_j d_j/\omega_j \geq 0,$$

against the alternative that it is strictly negative. The distribution of this statistics is simple N(0,1) and is equivalent to one of the tests described in Davidson and Duclos (2000). Once the $\gamma's$ are estimated together with their variance-covariance matrix, it is quite easy to find the standard deviation ω_j.

A.4.6 Testing for progressivity

Progressivity is tested by means of a series of inequalities. Let us define $K-1$ differences between normalized future and actual incomes $d_{i,i+1}$:

$$d_{i,i+1} = \frac{1}{y_i}\sum_{j=1}^{K} p_{ij}y_j - \frac{1}{y_{i+1}}\sum_{j=1}^{K} p_{i+1,j}y_j$$

that we want to be positive. We regroup these numbers in a vector d of dimension $K-1$. We have the following possible alternative hypothesis:

1. $H_0 : d = 0$
2. $H_1 : d \geq 0$
3. H_2 : no restriction on d

Dardanoni and Forcina (1998) or Formby et al. (2004) consider two types of tests that have a complicated distribution. They test T_{01} which means testing H_0 against H_1 and T_{12} which means testing H_1 against H_2. As explained in Davidson and Duclos (2000), these tests have a complicated geometry and it is far easier to test T_{21} which means H_2 against H_1. In this case the test is no longer a simultaneous test but can be computed as

$$T_{21} = \min_i \left(\frac{d_{i,i+1}}{\omega_{i,i+1}} \right) \sim N(0,1),$$

where ω_{ii} is the ith diagonal element of Ω the variance covariance matrix of d. In d, the only random element is the estimated matrix P while the y vector corresponds to a predefined grid. The i^{th} row of P are supposed to be independently drawn from a multinomial distribution with variance of each element given by $p_{ij}(1-p_{ij})/n_i$ and the covariance between p_{ij} and p_{ik} is given by $-p_{ij}p_{ik}/n_i$. So that finally, the maximum likelihood estimator $\hat{P}_i$ of each ith row of P is distributed according to

$$\sqrt{n_i}(\hat{P}_i - P_i) \sim N(0, \Sigma_i),$$

with

$$\Sigma_i = \begin{bmatrix} p_{i1}(1-p_{i1})/n_i & \cdots & -p_{i1}p_{iK}/n_i \\ & \ddots & \\ -p_{iK}p_{i1}/n_i & \cdots & p_{iK}(1-p_{iK})/n_i \end{bmatrix}$$

Let us now compute the variance of each element of d.

$$\begin{aligned} \mathrm{Var}(d_{i,i+1}) = \omega_{ii}^2 &= \frac{1}{y_i^2}\sum_{j=1}^{K} y_j^2 \mathrm{Var}(p_{ij}) + \frac{2}{y_i^2}\sum_{j=1}^{K-1}\sum_{k=j+1}^{K} y_k y_j \mathbf{Cov}(p_{ij}, p_{ik}) + \\ & \frac{1}{y_{i+1}^2}\sum_{j=1}^{K} y_j^2 \mathrm{Var}(p_{i+1,j}) + \frac{2}{y_{i+1}^2}\sum_{j=1}^{K-1}\sum_{k=j+1}^{K} y_k y_j \mathbf{Cov}(p_{i+1,j}, p_{i+1,k}) \end{aligned} \tag{A.6}$$

because the rows P_i and P_{i+1} are independent. If we cannot reject the null that the row are not comparable against the alternative of progressivity with this dominance test, we can compute a progressivity index as the one of Reynolds and Smolensky (1977) which is based on comparing two Gini indices. See Benabou and Ok (2001b), section 4.

A.5 Chapter 5

A.6 Properties of the Transition Matrix

A.6.1 Mobility indices

A transition matrix P has K independent rows. Each row indicates the probability to change from status j to status k the next period and sums to 1. Overall mobility can be summarised using a Prais (1955) index,

$$M_P(P) = (K - \mathrm{Tr}(P))/(K-1).$$

$M_P(P) = 0$ is perfect immobility while $M_P(P) = 1$ is perfect mobility.

Stochastic dominance test (monotone test)

According to the definition of monotone transition given by Benabou and Ok (2001c) :

Definition 4. *A transition matrix is said to be monotone if each row j is dominated by its follower $j+1$, which means*

$$\sum_{k=1}^{l} p_{j+1,k} \leq \sum_{k=1}^{l} p_{j,k}, \qquad \forall j, k = 1, 2, \ldots, K-1,$$

a monotone transition matrix must satisfy the fact that the sum of category probabilities $\sum_{k=1}^{L} P_{jk}$ for row i is larger than $\sum_{k=1}^{L} P_{j+1,k}$ for all $L < K$. This property is tested by means of series of inequalities. Let us define $K-1$ differences between sums of probabilities $d_{j,j+1}$ given L:

$$d_{j,j+1}^{L} = \sum_{k=1}^{L} P_{jk} - \sum_{k=1}^{L} P_{j+1,k}$$

There is total $(K-1)^2$ inequalities to be tested which can be stored into a matrix with dimension $(K-1) \times (K-1)$. In order to have a monotone matrix, it must satisfy the all the $(K-1)^2$ are positive. Or, if we want to only focus on the typical rows, row $j+1$ is said to dominate previous row j if for all L, the signs of the differences are satisfied. we have the following possbile alternative hypothesis:

1. H_0: $d = 0$
2. H_1: $d \geq 0$
3. H_2: no restriction on d

Dardanoni (1993) consider two types of test that have a complicated distribution. They test T_{01} which means testing H_0 against H_1 and which T_12 which means testing H_1 against H_2. As pointed out in Davidson and Duclos (2000), it is much easier to test T_21 (H_2 against H_1). In this case the test is no longer a simultaneous test but can be measured by :

$$T_{21} = \min_{j,L}\left(\frac{d_{j,j+1}^{L}}{\omega_{j,j+1}^{L}}\right) \sim N(0,1)$$

where $\omega_{j,j+1}^{L}$ is the jth diagonal element of Ω, the variance covariance matrix of d. In d, the only random element is the estimated matrix P while the y vector corresponds to a pre-defined grid. The jth row of P are supposed to be independently drawn from a multinomial distribution with variance of each element given by $P_{jk}(1-P_{jk})/n_j$ and the covariance between P_{jk} and P_{jm} is given by $-P_{jk}P_{jm}/n_j$. So that finally, the maximum likelihood estimator $\hat{P}_j$ of each jth row of P is distributed according to:

$$\Omega_j^L = \begin{bmatrix} P_{j1}(1-P_{j1})/n_j & \cdots & -P_{j1}P_{jL}/n_j \\ & \ddots & \\ P_{jL} - P_{j1}/n_j & \cdots & P_{jL}(1-P_{jL})/n_j \end{bmatrix} \tag{A.7}$$

Then we compute the variance of each element of d given L.

$$\begin{aligned} Var(d^L_{j,j+1}) &= \sum_{k=1}^{L} Var(P_{jk}) + 2\sum_{k=1}^{L-1}\sum_{m=k+1}^{L} Cov(P_{jk}, P_{jm}) && \text{(A.8)} \\ &+ \sum_{k=1}^{L} Var(P_{j+1,k}) + 2\sum_{k=1}^{L-1}\sum_{m=k+1}^{L} Cov(P_{j+1,k}, P_{j+1,m}) && \text{(A.9)} \end{aligned}$$

A.6.2 Proof

If two vectors a and b with zero mean and variance of 1 who are in an inner product space, according to Cauchy-Schwarz inequality we shall have:

$$|\langle a, b\rangle| \leq \sqrt{\langle a, a\rangle\langle b, b\rangle} \quad \text{(A.10)}$$

so that:

$$-1 \leq \frac{\langle a, b\rangle}{\sqrt{\langle a, a\rangle\langle b, b\rangle}} = \rho_{ab} \leq 1 \quad \text{(A.11)}$$

Given the fact that the correlations ρ_{ab} and ρ_{bc} [2] are within 0 and 1 while the correlation ρ_{ac} is unknown, the problem can be solved by using the orthgonal decomposition. Since both vector a and c are correlated to vector b and their correlations are known, we can rewrite a and c as:

$$\begin{aligned} a &= \langle a, b\rangle b + O^a_b \\ c &= \langle c, b\rangle b + O^c_b \end{aligned}$$

where O^a_b is the orthogonal projection of vector a onto b. Then the correlaion between a and c can be written as:

$$\rho_{ac} = \langle a, c\rangle = \langle \rho_{ab} b + O^a_b, \rho_{bc} b + O^c_b\rangle = \rho_{ab}\rho_{bc} + \langle O^a_b, O^c_b\rangle \quad \text{(A.12)}$$

and because that:

$$-1 \leq \frac{\langle O^a_b, O^c_b\rangle}{\sqrt{\langle O^a_b, O^a_b\rangle\langle O^c_b, O^c_b\rangle}} \leq 1$$

and:

$$\begin{aligned} \langle a, a\rangle &= \langle \rho_{ab} b + O^a_b, \rho_{ab} b + O^a_b\rangle = \rho^2_{ab} + \langle O^a_b, O^a_b\rangle \\ &\Rightarrow \langle O^a_b, O^a_b\rangle = 1 - \rho^2_{ab} \end{aligned}$$

2. Vector c has also mean of zero and variane of 1

because variance of a is 1, we then have the following condition that:

$$-\sqrt{(1-\rho_{ab}^2)(1-\rho_{bc}^2)} \leq \langle O_b^a, O_b^c \rangle \leq \sqrt{(1-\rho_{ab}^2)(1-\rho_{bc}^2)} \tag{A.13}$$

Finally, by replacing Inequation (A.13) into Equation A.12 we have:

$$\rho_{ab}\rho_{bc} - \sqrt{(1-\rho_{ab}^2)(1-\rho_{bc}^2)} \leq \rho_{ac} \leq \rho_{ab}\rho_{bc} + \sqrt{(1-\rho_{ab}^2)(1-\rho_{bc}^2)}$$

A.6.3 RNG

Draw a random number π from a truncated standard normal distribution, for example from $f(\pi|a < \pi < b)$, I apply an inverse sampling approach:

1. First draw r^{th} random number ξ^r from uniform(0,1) distribution.
2. Define $\bar{\xi}^r = (1-\xi^r)\Phi(a) + \xi^r\Phi(b)$
3. Obtain $\pi = \Phi^{-1}(\bar{\xi}^r)$ which relies between a and b.

Notice that the random numbers are drawn once and kept (McFadden 1989) when parameters vary during the MLE process.

Bibliography

Afridi, F., Li, S., and Ren, Y. (2012). Social identity and inequality: The impact of china's hukou system. Discussion Paper 6417.

Alesina, A. and Angeletos, G.-M. (2005). Fairness and redistribution. *American Economic Review*, pages 960–980.

Alesina, A., Di Tella, R., and MacCulloch, R. (2004). Inequality and happiness: are Europeans and Americans different? *Journal of Public Economics*, 88(9-10):2009–2042.

Alesina, A. and Glaeser, E. L. (2004). *Fighting poverty in the US and Europe: A world of difference.* Oxford University Press.

Alesina, A. and La Ferrara, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6):897–931.

Alesina, A. F. and Giuliano, P. (2009). Preferences for redistribution. Technical report, National Bureau of Economic Research.

Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(1):1–30.

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, 28(1):89–110.

Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2:244–263.

Atkinson, A. B. (1983). *The Economics of Inequality.* Clarendon Press, Oxford.

Atkinson, A. B., Bourguignon, F., and Morrisson, C. (1992). *Empirical Studies of Earnings Mobility*, volume 52. Harwood Academic Publishers.

Bauwens, L., Lubrano, M., and Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models.* Oxford University Press.

Bean, C. and Papadakis, E. (1998). A comparison of mass attitudes towards the welfare state in different institutional regimes, 1985–1990. *International Journal of Public Opinion Research*, 10(3):211–236.

Benabou, R. and Ok, E. A. (2001a). Mobility as progressivity: Ranking income processes according to equality of opportunity. Technical report, National Bureau of Economic Research.

Benabou, R. and Ok, E. A. (2001b). Mobility as progressivity: Ranking income processes according to equality of opportunity. Working Paper 8431, National Bureau of Economic Research, Cambridge, MA.

Benabou, R. and Ok, E. A. (2001c). Social mobility and the demand for redistribution: The POUM hypothesis. *The Quarterly Journal of Economics*, 116(2):447–487.

Benabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics. *The Quarterly Journal of Economics*, 121(2):699–746.

Berndt, E. K., Hall, B. H., and Hall, R. E. (1974). Estimation and inference in nonlinear structural models. In *Annals of Economic and Social Measurement, Volume 3, number 4*, pages 103–116. NBER.

Blanchflower, D. and Oswald, A. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, 88(7-8):1359–1386.

Blanchflower, D. G. and Oswald, A. J. (2003). Does inequality reduce happiness? Evidence from the states of the USA from the 1970s to the 1990s. Technical report, Warwick University.

Blanchflower, D. G. and Oswald, A. J. (2008). Is well-being U-shaped over the life cycle? *Social Science and Medicine*, 66(8):1733–1749.

Bogenhold, D. and Staber, U. (1991). The decline and rise of self-employment. *Work, Employment and Society*, 5(2):223–239.

Boudon, R. (1973). *Mathematical Structures of Social Mobility.* Elsevier Scientific Pub. Co., Amsterdam and New York.

Bound, J. and Johnson, G. (1992). Changes in the structure of wages in the 1980's: An evaluation of alternative explanations. *American Economic Review*, 82(3):371–92.

Bourguignon, F. (2012). *La mondialisation de l'inégalité.* Seuil, Paris.

Boyce, C. J. (2010). Understanding fixed effects in human well-being. *Journal of Economic Psychology*, 31(1):1–16.

Boyce, C. J., Brown, G. D., and Moore, S. C. (2010). Money and happiness : Rank of income, not income, affects life satisfaction. *Psychological Science*, 21(4)(4):471–475.

Butler, J. and Moffitt, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50:761–764.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, 16(5):1190–1208.

Cappellari, L. and Jenkins, S. P. (2003). Multivariate probit regression using simulated maximum likelihood. *The Stata Journal*, 3(3):278–294.

Cappelli, P. and Sherer, P. D. (1988). Satisfaction, market wages, and labor relations: An airline study. *Industrial relations*, 27:57–73.

Chang, S.-K. (2011a). Computationally practical simulation estimation for dynamic panel tobit models. *Academia Economic Papers*, 39(1):1–32.

Chang, S.-K. (2011b). Simulation estimation of two-tiered dynamic panel Tobit models with an application to the labor supply of married women. *Journal of Applied Econometrics*, 26(5):854–871.

Chen, S. and Ravallion, M. (2008). The developing world is poorer than we thought, but no less successful in the fight against poverty. Policy Research Working Paper 4703, World Bank, Washington, DC.

Chen, Y. and Cowell, F. A. (2013). Mobility in china. Technical report, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Clark, A. (1996). L'utilité est-elle relative? analyse à l'aide de données sur les ménages. *Economie et Prévision*, 121:151–164.

Clark, A. and d'Angelo, E. (2008a). Upward social mobility, well-being and political preferences: Evidence from the BHPS. Technical report, Paris School of Economics.

Clark, A. and d'Angelo, E. (2008b). Upward social mobility, well-being and political preferences: Evidence from the BHPS. Working paper, Paris School of Economics, PSE, Paris.

Clark, A., Flèche, S., and Senik, C. (2014). Economic growth evens-out happiness: Evidence from six surveys. Working Paper 2014-032, Paris School of Economics, Paris.

Clark, A. and Oswald, A. (1996). Satisfaction and comparison income. *Journal of Public Economics*, 61(3):359–381.

Clark, A. and Senik, C. (2011). Will GDP growth increase happiness in developing countries? Discussion Paper 5595, IZA.

Clark, A. E. (2003). Inequality-aversion and income mobility: A direct test. Technical report, Delta.

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1):95–144.

Clark, A. E. and Georgellis, Y. (2013). Back to baseline in Britain: Adaptation in the British Household Panel Survey. *Economica*, 80(319):496–512.

Conlisk, J. (1990). Monotone mobility matrices. *The Journal of Mathematical Sociology*, 15(3-4):173–191.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5):829–844.

Cramer, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, N.J.

Crouchley, R. (1995). A random-effects model for ordered categorical data. *Journal of the American Statistical Association*, 90(430):489–498.

Dardanoni, V. (1993). On measuring social mobility. *Journal of Economic Theory*, 61:372–394.

Dardanoni, V. (1995). Income distribution dynamics: monotone Markov chains make light work. *Social Choice and Welfare*, 12:181–192.

Dardanoni, V. and Forcina, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *Journal of the American Statistical Association*, 93(443):1112–1123.

Davidson, R. and Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, 68(6):1435–1464.

Deleeck, H. (1989). The adequacy of the social security system in Belgium, 1976-1985. *Journal of Social Policy*, 18(1):91–117.

Demurger, S., Sachs, J. D., Woo, W. T., Bao, S., and Chang, G. (2002). The relative contributions of location and preferential policies in China's regional development: Being in the right place and having the right incentives. *China Economic Review*, 13(4):444–465.

Diener, E., Oishi, S., and Lucas, R. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual Review of Psychology*, 54(1):403–425.

Doan, T. A. (2010). *RATS version 8 : user's guide.* Estima.

Easterlin, Richard, A. (1974). Does economic growth improve the human lot? In David, P. A. and Reder, Melvin, W., editors, *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz*. Academic Press, Inc., New York.

Edlund, J. (2008). Nominal category class classifications. CESSDA PPP workpackage 9th meeting held in Paris, France.

Egger, P., Pfaffermayer, M., and Weber, A. (2007). Sectoral adjustment of employment to shifts in outsourcing and trade: Evidence from a dynamic fixed effects multinomial logit model. *Journal of Applied Econometrics*, 22(3):559–580.

Fang, C., Zhang, X., and Fan, S. (2002). Emergence of urban poverty and inequality in China: Evidence from household survey. *China Economic Review*, 13(4):430–443.

Feldman, S. and Zaller, J. (1992). The political culture of ambivalence: Ideological responses to the welfare state. *American Journal of Political Science*, pages 268–307.

Feri, F. (2012). A note on the POUM effect with heterogeneous social mobility. *Economics Letters*, 115:258–262.

Ferrer-i-Carbonell, A. (2005). Income and well-being: An empirical analysis of the comparison income effect. *Journal of Public Economics*, 89(5-6):997–1019.

Ferrer-i-Carbonell, A. and Frijters, P. (2004). How important is the methodology for the estimates of the determinants of happiness. *The Economic Journal*, 114(497):641–459.

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82(2):225–246.

Formby, J. P., Smith, W. J., and Zheng, B. (2004). Mobility measurement,transition matrices and statistical inference. *Journal of Econometrics*, 120:181–205.

Forsé, M. and Parodi, M. (2007). Perception des inégalités économiques et sentiment de justice sociale. *Revue de l'OFCE*, 102:483–540.

Frey, B. S. and Stutzer, A. (2002). What can economists learn from happiness research. *Journal of Economic Literature*, 40(2):402–435.

Gabaix, X. and Landier, A. (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics*, 123(1):49–100.

Ganzeboom, H. B., Luijkx, R., and Treiman, D. J. (1989). Inter-generational class mobility in comparative perspective. *Research in social Stratification and Mobility*, 8:3–84.

Gershon, L. (2014). Self-employment: Exciting opportunity or last resort?

Goedhart, T., Halberstadt, V., Kapteyn, A., and van Praag, B. (1977). The poverty line: Concept and measurement. *Journal of Human Resources*, 12(4):503–520.

Greb, F., Klasen, S., Pasaribu, S., and Wiesenfarth, M. (2011). Dollar a day re-revisited. Discussion Paper 91, Georg-August-Universitat, Gottingen.

Guilbaud, O. (1977). Estimating the probability eigenvector and related characteristics of an ergodic transition matrix. *Scandinavian Journal of Statistics*, 4:105–112.

Guillaud, E. (2013). Preferences for redistribution: an empirical analysis over 33 countries. *Journal of Economic Inequality*, 11:57–78.

Hajivassiliou, V., Savignac, F., and de France, B. (2011). Novel approaches to coherency conditions in LDV models with an application to interactions between financing constraints and a firm's decision and ability to innovate. *LSE discussion papers.*

Hansen, B. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.

Hart, P. E. (1976). The comparative statics and dynamics of income distributions. *Journal of the Royal Statistical Society. Series A (General)*, 139(1):108–125.

Heckman, J. (1981a). The incidental parameter problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Panel Data with Econometric Applications*, chapter 4, pages 179–195. MIT press.

Heckman, J. (1981b). Statistical models for discrete panel data. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Panel Data with Econometric Applications*, chapter 3, pages 114–178. MIT press.

Helliwell, J. F. and Wang, S. (2012). The state of world happiness. *World Happiness Report*, pages 10–57.

Hendrickx, J. (2000). Special restrictions in multinomial logistic regression. *Stata Technical Bulletin*, 56:18–26.

Hirschman, A. and Rothschild, M. (1973). The changing tolerance for income inequality in the course of economic development, with a mathematical appendix. *Quarterly Journal of Economics*, 87:544–566.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.

Hopkins, D. J. and King, G. (2010). Improving anchoring vignettes designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*, 74(2):201–222.

Houkai, W. and Xiaoxia, W. (2009). The review and prospect of china's anti-poverty policies [j]. *The Journal of Shanghai Administration Institute*, 2:008.

Hungerford, T. L. (1993). U.S. income mobility in the seventies and eighties. *Review of Income and Wealth*, 31(4):403–417.

Idema, T. and Rueda, D. (2011). Redistribution preferences and life-cycle income. Technical report, University of Oxford.

Jenkins, S. P. (2000). Modelling household income dynamics. *Journal of population economics*, 13(4):529–567.

Jenkins, S. P. (2010). The british household panel survey and its income data. Technical report, Discussion paper series//Forschungsinstitut zur Zukunft der Arbeit.

Kahneman, D., Diener, E., and Schwarz, N. (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.

Kanbur, R. and Zhang, X. (1999). Which regional inequality? The evolution of rural–urban and inland–coastal inequality in China from 1983 to 1995. *Journal of Comparative Economics*, 27(4):686–701.

Kapteyn, A., Kooreman, P., and Willemse, R. (1988). Some methodological issues in the implementation of subjective poverty definitions. *Journal of Human Resources*, 23(2):222–242.

Karabarbounis, L. (2011). One dollar, one vote. *The Economic Journal*, 121(553):621–651.

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics*, 107:35–78.

Khan, A. R. and Riskin, C. (1998). Income and inequality in China: composition, distribution and growth of household income, 1988 to 1995. *The China Quarterly*, 154:221–253.

Khan, A. R. and Riskin, C. (2002). Inequality and poverty in China in the age of globalization. *The China Quarterly*, 169:212–214. Special Issue: China and Europe since 1978: A European Perspective.

King, G., Murray, C. J., Salomon, J. A., and Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1):191–207.

Knight, J., Song, L., and Gunatilaka, R. (2009). Subjective well-being and its determinants in rural China. *China Economic Review*, 20(4):635–649.

Larsen, R. J., Diener, E., and Emmons, R. A. (1985). An evaluation of subjective well-being measures. *Social Indicators Research*, 17(1):1–17.

Levy, H. (1973). Stochastic dominance among log-normal prospects. *International Economic Review*, 14(3):601–614.

Lyubomirsky, S. (2011). Hedonic adaptation to positive and negative experiences. *Oxford handbook of stress, health, and coping*, pages 200–224.

McBride, M. (2001). Relative-income effects on subjective well-being in the cross-section. *Journal of Ecnomic Behavior and organization*, 45:251–178.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *econometrica*, 57(5):995–1026.

Melenberg, B. (1992). Micro-econometric models of consumer behavior and welfare.

Meltzer, A. H. and Richard, S. F. (1981). A rational theory of the size of government. *The Journal of Political Economy*, 89(5):914–27.

Muller, W. (2000). Casmin educational classification. Technical report, Nuffield College, Oxford University.

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

Murphy, K. and Welch, F. (1992). The structure of wages. *Quarterly Journal of Economics*, 107(1):285–326.

Neustadt, I. (2011). Do religious beliefs explain preferences for income redistribution? experimental evidence. *CESifo Economic Studies*, 57(4):623–652.

OECD (2011). An overview of growing income inequalities in OECD countries: Main findings. OECD database on household income distribution and poverty, OECD.

Osberg, L. and Sharpe, A. (2002). Comparisons of trends in GDP and economic well-being - the impact of social capital. Discussion paper, OECD, Paris.

Oswald, A. J., Proto, E., and Sgroi, D. (2009). Happiness and productivity. Technical report, IZA discussion papers.

Peng, K., Nisbett, R. E., and Wong, N. Y. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods*, 2(4):329.

Piketty, T. (1995). Social mobility and redistributive politics. *The Quarterly Journal of Economics*, 110(3):551–584.

Pradhan, M. and Ravallion, M. (2000). Measuring poverty using qualitative perceptions of consumption adequacy. *The Review of Economics and Statistics*, 82(3):462–471.

Prais, S. J. (1955). Measuring social mobility. *Journal of the Royal Statistical Society, Series A, Part I*, 118(1):56–66.

Rainer, H. and Siedler, T. (2008). Subjective income and employment expectations and preferences for redistribution. *Economics Letters*, 99(3):449–453.

Ravallion, M. (1994). *Poverty Comparisons*. Harwood Academic Press, Chur, Switzerland.

Ravallion, M. and Chen, S. (2001). How did the world's poorest fared in the 1990s. *Review of Income and Wealth*, 47(3):283–300.

Ravallion, M. and Chen, S. (2004). How did the world's poorest fared since the early 1980s. *World Bank Reserch Observer*, 19(2):141–170.

Ravallion, M., Chen, S., and Sangraula, P. (2009). Dollar a day revisited. *The World Bank Economic Review*, 23(2):163–184.

Ravallion, M., Datt, G., and van de Walle, D. (1991). Quantifying absolute poverty in the developing world. *Review of Income and Wealth*, 37(4):345–361.

Ravallion, M. and Lokshin, M. (2000). Who wants to redistribute?: The tunnel effect in 1990s Russia. *Journal of public Economics*, 76(1):87–104.

Rawls, J. (1971). *A Theory of Justice*. Oxford University Press, London.

Rendon, S. R. (2012). Fixed and random effects in classical and Bayesian regression. *Oxford Bulletin of Economics and Statistics*. Forthcoming.

Reynolds, M. and Smolensky, E. (1977). *Public Expenditures, Taxes and the Distribution of Income: The United States 1950, 1961, 1970*. Academic Press, New York.

Rodrik, D. (1998). Why do more open economies have bigger governments. *Journal of Political Economy*, 106(5):997–1032.

Romer, T. (1975). Individual welfare, majority voting and the properties of a linear income tax. *Journal of Public Economics*, 7:163–188.

Sandvik, E., Diener, E., and Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non-self-report measures. *Journal of Personality*, 61(3):317–342.

Scheier, M. F. and Carver, C. S. (1992). Effect of optimism on psychological and physical well-being: Theoretical overview and empirical update. *Cognitive Therapy and Research*, 16(2):201–228.

Scheve, K. and Stasavage, D. (2006). Religion and preferences for social insurance. *Quarterly Journal of Political Science*, 1(3):255–286.

Schwarze, J. and Harpfer, M. (2007). Are people inequality averse, and do they prefer redistribution by the state? Evidence from German longitudinal data on life satisfaction. *Journal of Socio-Economics*, 36(2):233–249.

Sen, A. (1980). *Equality of what?*, volume 1. na.

Sen, A. (1993). Capability and well-being. In Nussbaum, M. and Sen, A., editors, *The Quality of Life*. Clarendon Press, Oxford.

Sen, A. (1999). Commodities and capabilities. *OUP Catalogue*.

Senik, C. (2005). Income distribution and well-being: what can we learn from subjective data? *Journal of Economic Surveys*, 19(1):43–63.

Shane, F. and Loewenstein, G. (1999). Hedonic adaptation. In Kahneman, D., Diener, E., and Schwarz, N., editors, *Well-Being: The Foundations of Hedonic Psychology*, pages 302–329. Russell Sage Foundation, New York.

Sherlaw-Johnson, C., Gallivan, S., and Burridge, J. (1995). Estimating a Markov transition matrix from observational data. *The Journal of the Operational Research Society*, 46(3):405–410.

Shorrocks, A. F. (1976). Income mobility and the markov assumption. *The Economic Journal*, 86(343):566–578.

Stevenson, B. and Wolfers, J. (2008). Economic growth and subjective well-being: Reassessing the easterlin paradox. Technical report, National Bureau of Economic Research.

Stewart, M. (2006). Maximum simulated likelihood estimation of random-effects dynamic probit models with autocorrelated errors. *The Stata Journal*, 6(2):256–272.

Stewart, M. B. (2004). Semi-nonparametric estimation of extended ordered probit models. *The STATA Journal*, 4:27–39.

Tanner, M. S. (2006). We the people (of China)... *Wall Street Journal*. February 2.

Tendulkar, S. D. (2009). Report of the expert group to review the methodology for estimation of poverty. Technical report, Government of India Planning Commission, New Dehli.

Thurow, L. (1971). The income distribution as a pure public good. *The Quarterly Journal of Economics*, 85(2):327–336.

Tuma, N. B. and Robins, P. K. (1980). A dynamic model of employment behavior: An application to the Seattle and Denver income maintenance experiments. *Econometrica*, 48(4):1031–1052.

Van den Bosch, K., Callan, T., Estivill, J., Hausman, P., Jeandidier, B., Muffeis, R., and Yfantopoulos, J. (1993). A comparison of poverty in seven European countries and regions using subjective and relative measures. *Journal of Population Economics*, 6(3):235–259.

van Praag, B. (1971). The welfare function of income in Belgium: An empirical investigation. *European Economic Review*, 2(3):337–369.

van Praag, B. and Ferrer-i-Carbonell, A. (2004). *Happiness Quantified: A Satisfaction Calculus Approach Oxford*. Oxford University Press, Oxford: UK.

Van Praag, B. M. and Ferrer-i-Carbonell, A. (2006). An almost integration-free approach to ordered response models. Technical report, Tinbergen Institute, Faculty of Economics and Econometrics, University of Amsterdam.

Walter, S. (2010). Globalization and the welfare state: Testing the microfoundations of the compensation hypothesis. *International Studies Quarterly*, 54(2):403–426.

Whyte, M. K., Maocan, G., Kwok-bun, C., Ku, A. S., and Yin-wah, C. (2009). How angry are Chinese citizens about current inequalities? Evidence from a National Survey. *Social Stratification in Chinese Societies*, page 17.

Wong, K., Fu, D., Li, C. Y., and Song, H. X. (2007). Rural migrant workers in urban China: living a marginalised life. *International Journal of Social Welfare*, 16(1):32–40.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

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

World bank (2008). Global purchasing power parities and real and real expenditure 2005. Technical report, World Bank, Washington DC.

Wu, X. (2007). The Chinese household registration and intergenerational social mobility. *Sociological Studies*, 6:38–65.

Xu, J. and Liu, H. (2013). Social justice recognition, fluidity expectation and the preference of residence redistribution-an empirical study based on CGSS data. *Journal of Yunnan University of Finance and Economics*, 2.

Xun, Z. and Lubrano, M. (2014). Simulation estimation of two-tiered dynamic panel tobit models with an application to the labor supply of married women: A comment. Technical report, GREQAM, Aix-Marseille University.