Topics in distributional analysis - The importance of intermediate institutions for income distributions, inequality and intra-distributional mobility

For the degree of Ph.D.

London School of Economics

Christian Schluter

January 1997

Contents

1 Introduction x

1.1 Part I -Chapter 2: The evolution of wealth distributions and the economics of information xv

1.2 Chapter 3: On the performance of social benefit systems xvi

1.3 Part II - Chapter 4: The statistical analysis of inequality and mobility indices xviii

1.4 Part III - Chapter 5: The panel dataset GSOEP described xix

1.5 Chapter 6 : Income distribution and inequality in Germany xx

1.6 Chapter 7 : Income dynamics in Germany xxii

1.7 Chapter 8 : On the non-stationarity of German income mobility (and some observations about poverty) xxiv

2 The evolution of wealth distributions and the economics of information 2

2.1 Introduction 3

2.1.1 The problem in context 4
2.2 The static partial equilibrium model 8
2.2.1 The incentive problem in a partial equilibrium model 15
2.3 The general equilibrium model 18
2.3.1 Decreasing absolute risk aversion (DARA) . . 20
2.3.2 The simulation and some comparative statics 27
2.3.3 Constant absolute risk aversion (CARA) . . . 34
2.4 Dynamically evolving wealth distributions 35
2.4.1 Parameter values and the nature of the dynamics 37
2.5 Conclusions 41
2.6 Appendix: the ergodic theorem and the well-behaved case 43

3 On the performance of social benefit systems 46
3.1 Introduction 47
3.1.1 The context of the problem 49
3.2 The model 51
3.3 Evaluating the performance of the benefit regimes . . 63
3.3.1 Poverty 64
3.3.2 Inequality 66
3.3.3 Social welfare and work incentives: a trade-off ? 67
3.4 Conclusion 70
3.5 Appendix: Risk aversion, tax-financed benefits, and Pareto improvements 72

4 **The statistical analysis of inequality and mobility indices** **76**

4.1 Introduction 77

4.2 Asymptotic methods when the empirical distribution function is used only once. 79

4.2.1 Asymptotic distributions of inequality and mobility measures 79

4.2.2 A framework for hypothesis testing 85

4.3 Resampling Plans 91

4.3.1 The smooth bootstrap estimators 93

4.3.2 A simulation: Confidence intervals 94

4.3.3 Discussion 96

4.4 Robustness 98

4.4.1 Robustness properties of quantile based inequality measures 101

4.4.2 Decomposable inequality measures 102

4.4.3 Remedies 103

4.5 Conclusions 104

5 **The Panel Dataset GSOEP Described** **105**

5.1 Introduction 106

5.2 The panel: description of the raw data 106

5.3 The derived data, simulations, and income definitions 109

5.4 Representativeness 113

6 **Income Distribution and Inequality in Germany** **116**

6.1 Introduction 117

6.2 Estimating densities and assessing their properties . . 119

6.2.1 Kernel density estimation 119

6.2.2 A bootstrap test for multimodality 121

6.3 The estimates: The distributional consequences of the German tax-benefit system 124

6.4 Explaining income inequality: The anatomy of inequality in 1991 129

6.4.1 The German income distribution in context . 140

6.5 Conclusions 142

7 **Income Dynamics in Germany** **145**

7.1 Introduction 146

7.2 Cross sectional dynamics: a changing profile of inequality in the 90s? 149

7.3 Intra-distributional mobility 155

7.4 Conclusions 169

8 **On the Non-Stationarity of German Income Mobility (and some observations on poverty dynamics)** **171**

8.1 Introduction 172

8.2 The data 178

8.3 Pure Markov models in discrete time 180

8.3.1 First order Markov chains 180

8.3.2 Second order Markov chains 183

8.3.3 First order against second order Markov chains 183
8.4 Mixed Markovian models in discrete time 184
8.4.1 A mover-stayer model : the stationary case . . 185
8.4.2 A mover-stayer model: the non-stationary case 189
8.5 Poverty re-examined 191
8.5.1 A Markov model with observed heterogeneity 191
8.5.2 Semi-Markov processes: non-stationary duration models 196
8.5.3 German poverty dynamics in context 204
8.6 Conclusion 207
8.7 Appendix: The EM-Algorithm 209

to my parents

Acknowledgements

Many people have suffered through the innumerable drafts of the papers contained in this dissertation. My greatest debt is to my supervisor Frank Cowell on whom the heaviest burden fell. I am also grateful to Ramses Abul-Naga, Tony Atkinson, Tim Besley, Jean-Yves Duclos, Chico Ferreira, Stephen Jenkins, John Micklewright, Danny Quah, and Maria-Pia Victoria-Feser for comments and encouragement.

Most of the material presented here has been circulated in one form or another. Chapter 3 is forthcoming in the Economic Journal, vol. 107, March 1997. Papers based on earlier drafts of chapters 6 and 7 will appear in Proceedings of the 1996 conference of the German Association of Political Economy, "Aspects of the distribution of income", Metropolis Verlag Marburg, in January 1997, and in Proceedings of the 1996 international conference of the German Socio-Economic Panel Study users, DIW Vierteljahreshefte zur Wirtschaftsforschung in June 1997. Earlier drafts of chapters 6, 7, 3, 8 have appeared as LSE distributional analysis research programme discussion papers 16, 17, 28, and 30. The last has also appeared as a Working Paper at the European University Institute, Florence.

I am also grateful to STICERD for providing a home for me at the LSE.

Abstract

The unifying theme of this dissertation is the importance of intermediate institutions for income distributions, inequality and intra-distributional mobility. First, we analyse the effects of informational problems in a general equilibrium model with dynamically evolving wealth distributions. Poor agents need to borrow funds but a non-commitment problem on the capital market leads to persistent inequality. The next important institution to be examined is the tax-benefit system. The third chapter investigates the relative performance of alternative unemployment benefit regimes in a search-theoretic general equilibrium model of the labour market. Policy objectives such as the reduction in inequality or the alleviation of poverty are considered and the incentive problems are examined. Prior to the empirical analysis, the fourth chapter develops the large sample distribution of a number of inequality and mobility indices. Moreover, the relative performance of these (asymptotic) approximations and various bootstrap estimators is examined. The data is described in chapter five. The sixth chapter analyses the distributional consequences of the German tax-benefit system using the German Socio-Economic Panel. Two dimensions income dynamics are investigated by distinguishing between shape dynamics and intra-distributional mobility. The complementarity between various tools such as non-parametric stochastic kernel density estimates and transition matrices is explored. As the transition probabilities are found to be time-varying, several statistical models of income mobility are

estimated (and a new mover-stayer model is proposed) in the last chapter. In order to give an economic explanation of the observed mobility patterns various duration models (with duration dependent hazards and unobserved heterogeneity) are estimated.

Chapter 1

Introduction

This dissertation contributes to the explanation of income and wealth distributions, and their evolutions in time. It is obvious that no single universal explanation can be given for such a complex phenomenon. As Champernowne (1952) observes:

> "The forces determining the distribution of incomes in any community are so varied and complex, and interact and fluctuate so continuously, that any theoretical model must either be unrealistically simplified or hopelessly complicated. We shall choose the former alternative but then give indications that the introduction of some of the more obvious complications of the real world does not seem to disturb the general trend of our conclusions." (p.246)

In the same vein, Atkinson (1994), "(s)eeking to explain the distribution of incomes", emphasises time and again that the distribution of incomes amongst persons cannot be identified with the distribution of factor incomes; distributional analysis needs to go beyond the latter. To illustrate this he singles out diverse and complex issues which need to be taken into account such as: differences (inhomogeneities) of incomes, human capital, diversity of income sources, intervening institutions such as social insurance, income from abroad, and the influence of the state budget. In an attempt to illustrate the many inter-relationships between them, he presents a diagram involving 25 arrows. Moreover, all these aspects are overlapping, complementary, and not mutually exclusive. They constitute economic forces, unfolding in time, which lead, individually, to greater or lesser inequality. Given this complexity, particular aspects have to be singled out, and the chosen elements have to be examined in isolation.

However, the observation that the theory of personal income distribution extends beyond the scope of the theory of the distribution of factor incomes had already been made by Dalton (1920) in his inquiry into "Some aspects of the inequality of incomes in modern communities". He wrote:

> "While studying economics at Cambridge in 1909-10, I became specially interested in those [parts] which set out to discuss the distribution of income. I gradually noticed, however, that most 'theories of distribution' were

> almost wholly concerned with distribution between 'factors of production'. Distribution as between persons, a problem of more direct and obvious interest, was either left out of the textbooks altogether, or treated so briefly, as to suggest that it raised no question, which could not be answered either by generalisations about the factors of production, or by plodding statistical investigations, which professors of economics were content to leave to lesser men." (p. vii)[1]

Before summarising the thesis in detail, it is perhaps useful to describe the intellectual space spanned by its eight chapters. This dissertation focuses on the importance of intermediate or "intervening institutions" (Atkinson (1994)) for the distributional outcome. Such intermediate institutions are many, and the dissertation singles out just two of them. First, chapter 2 analyses the distributional consequences of informational problems in a dynamic general equilibrium model in which the capital market is modelled explicitly. The behaviour of financial intermediaries confronted by such problems turns out to be a powerful mechanism for making inequality

[1]The last sentence is reminiscent of William Playfair's observation made in the Statistical Breviary (1801): ".. there is no study less alluring or more dry and tedious than statistics, unless the mind and the imagination are set to work or that the person is particularly interested in the subject; which is seldom the case with young men in any rank in life." However, by applying our statistical tools in the empirical chapters of this dissertation we hope to set mind and imagination to work and tell interesting and important stories about income dynamics.

persistent. A quite different type of intermediate institution is the tax-benefit system[2]. It is an institution which is particularly prominent in developed economies since public transfers constitute, after earnings, the second most important component of income. Chapter 3 examines the performance of alternative benefit systems in a search-theoretic general equilibrium model of the labour market. The relative performance of flat-rate and earnings-related unemployment benefits is assessed in the light of policy objectives such as the reduction of inequality or the alleviation of poverty. Although the point of departure of this chapter is an empirical observation, the analysis is conducted entirely in the context of a theoretical model.

By contrast, chapters 5 to 8 present an essentially empirical analysis for "plodding statistical investigations" is an integral part of distributional analysis (and one approach pursued here is plotting non-parametric estimates of the income distribution). After all, one of the principal concerns of distributional analysis is the question 'What does the distribution look like?' spawning other questions such as 'Has its shape changed? ' and 'How many modes are there and where are they? ' In these chapters the distributional consequences of the German tax-benefit system are examined. At first, the focus of chapter 6 is one particular year. Chapters 7 and 8, however, analyse the dynamics of the post-tax post-benefit income

[2]As Atkinson (1996) observes: "The gross incomes generated by production are typically modified by taxation, used to finance public spending, including transfers which constitute [another] source of personal incomes."

distribution. The use of a panel dataset is especially powerful in this context since one can go beyond the analysis of moving cross-sections and investigate the other dimension of income dynamics. Chapter 7 divides the analysis of income dynamics into its two constituent components by distinguishing between the shape dynamics and intra-distributional mobility. Whilst chapters 6 and 7 are exploratory and discuss the merits of various analytical tools, chapter 8 attempts to explain the observed income dynamics. First, the non-stationarity of the data is analysed and various statistical models are estimated. Then the focus is one important income state - poverty - instead of the entire transition matrix. Questions addressed are about the incidence and duration of poverty, and the extent to which the tax-benefit system succeeds or fails in alleviating poverty.

Despite its conceptual unity, this dissertation is organised as a collection of essays, which are grouped into three parts. In Part I, the above issues are analysed in a theoretical fashion as the analysis is conducted in the context of specific models. Part II develops statistical tools and collects results about some statistical aspects of distributional analysis, such as the asymptotic distributions of certain inequality and mobility indices. Various useful statistical techniques are described and discussed, such as kernel density estimation, bootstrapping, and tests for multi-modality. These statistical tools are then applied in the empirical chapters of Part III, in which a panel dataset, the German Socio-Economic Panel (GSOEP), is employed.

A more detailed outline of the issues tackled in the separate chapters is as follows.

1.1 Part I -Chapter 2: The evolution of wealth distributions and the economics of information

In this chapter we examine the distributional consequences of informational problems in the context of a dynamic general equilibrium model in which capital markets are modelled explicitly. It turns out that the economics of information give rise to powerful mechanisms which lead to persistent inequality. This model thus isolates a new economic channel because standard models, which invoke perfect markets, typically predict the vanishing of inequality. Far from converging to a mass point, the distributional outcome is a nondegenerate distribution.

In order to concentrate on this new economic channel, agents are assumed to be identical except for their inherited wealth. They can be productive only if they undertake a project. However, any project is costly and requires the payment of an indivisible admission fee or initial factor outlay. Agents who cannot afford this try to borrow on the capital market. Banks offer a financial contract which specifies an interest rate and a collateral. An incentive problem emerges from the incompleteness of the financial contract: borrowers cannot commit not to undertake a risky activity (which has a lower expected return for banks). But banks can indirectly control the type of project undertaken by appropriate contract design. We analyse the static equilibrium of the model and trace the dynamic consequence of this informational problem.

1.2 Chapter 3: On the performance of social benefit systems

The motivation for this chapter is the observation that social insurance is an important institution which 'stands between' the distribution of factor incomes and the final empirical (post-tax post-benefit) distribution of disposable incomes (recall Atkinson's observation cited in footnote 2). It is a prominent institution in developed economies, designed to protect economic agents against income risk. The form of social insurance schemes found in practice varies considerably: unemployment benefit in the UK is a flat rate, whereas it is earnings related on the Continent and the US; the replacement rates and the methods of finance differ. This chapter analyses the performance of alternative unemployment benefit systems in a search-theoretic framework. The relative performance of flat-rate and earnings-related unemployment benefits will be assessed in the light of policy objectives such as the reduction in inequality and the alleviation of poverty. However, these policy objectives may not command a universal consensus because of either a diversity of opinion or an intrinsic arbitrariness in the parameters characterising the social welfare function. So we might ask whether the ranking of the benefit regimes depends on the parameters of the social objective; although people may disagree about parameters, could they agree on a ranking ?

There is also a potential trade-off between the equity objective of poverty alleviation and the efficiency consideration of work incentives. A greater benefit might increase the income of an unemployed beneficiary and thereby reduce the difference between his income and the poverty line. For a person who remains in poverty, this increase reduces his poverty. But an increase in benefits reduces the incentives to work (particularly for those persons with a current low job productivity). Consequently, unemployment might rise, possibly increasing the numbers of the poor, raising aggregate poverty. This potential trade-off is examined in a general equilibrium setting. Again, does the resolution of this trade-off depend on the parameters of the social objective, or is it unambiguous for all admissible parameters ?

1.3 Part II - Chapter 4: The statistical analysis of inequality and mobility indices

This chapter constitutes the second part of the dissertation. By focusing exclusively on the (often neglected) statistical problems in distributional analysis we establish an indispensable statistical framework to which the subsequent empirical investigations have recourse in order to make statistically rigorous statements.

We summarise results on the asymptotic distributions of standard inequality measures and derive a framework for drawing inferences. A novel result is the derivation of the asymptotic distribution of the Shorrock (1978) mobility index.

Instead of using the (Gaussian) approximations, an alternative approach is to bootstrap the test statistics. It is thus natural to investigate the relative performance of these two techniques for sample sizes typically encountered in empirical work. The approach pursued here is to compare the lengths of the various confidence intervals. Finally the robustness properties of the inequality indices are examined.

1.4 Part III - Chapter 5: The panel dataset GSOEP described

This chapter describes the German Socio-Economic Panel (GSOEP) used in the subsequent empirical part of the dissertation.

These later chapters seek to assess the welfare properties of the income distribution and the distributional consequences of the German tax-benefit system. To this end, data on disposable income -the best monetary measure of well-being- is needed. Unfortunately, GSOEP is most user-unfriendly in this respect, since no reliable raw data on net income is supplied. However, there is an extensive supply of information on income sources and pre-tax incomes. There is no alternative but to estimate post-tax post-benefit income through a tax-benefit simulation. This task is equally unglororius but equally essential as of Johnson (1755)'s compiler of dictionaries about whom he observed:

> "It is the fate of those who dwell at the lower employments of life, to be rather driven by the fear of evil, than attracted by the prospects of good; to be exposed to censure, without hope of praise; to be disgraced by miscarriage, or punished for neglect, where success would have been without applause, and diligence without rewards.
>
> Among these unhappy mortals is the writer of dictionaries."

Compilers of dictionaries and datasets appear to share a similar fate.

This chapter discusses various data issues, income definitions, and the tax-benefit simulations. Since one aim of the empirical investigation is to make inferences about the (West German) population, the issue of representativeness of the sample (sample selection and weighting procedures) is discussed.

1.5 Chapter 6 : Income distribution and inequality in Germany

The focus of this empirical chapter is the German tax-benefit system and its distributional consequences. Defering the analysis of income dynamics to subsequent chapters, only the 1991 cross-section taken from GSOEP is examined in detail.

In the second part of this chapter we examine the anatomy of that income inequality which remains after the tax-benefit system has modified the distribution. To what extent do personal characteristics influence the shape of the resulting distribution ? Decomposition analyses show that the overall distribution of post-tax post-benefit income is a mixture of various underlying distributions. The low income groups are particularly dominated by foreign nationals, a sizeable group of the elderly, and those out of work.

These issues are addressed using two complementary approaches.

In the first step, in order to assess the welfare properties of the income distribution and to make statistical inferences, conventional tools such as inequality indices and Lorenz curves are applied. For instance, we show that the net income distribution Lorenz dominates (statistically significantly) the gross income distribution.

In the second step, the non-parametric technique of kernel density estimation is employed. The shape of the income distribution is thus captured directly, instead being infered indirectly from a set of estimated shape parameters of a parametric model or some other summary statistics. We examine 'graphical' features of particular interest (not only by inspection but also using statistical procedures) such as the number and the location of the modes of the distribution. Although this technique has its own problems, it avoids some heroic distributional assumptions of parametric methods, and is a natural tool for an exploratory analysis driven by the question 'What does the income distribution look like? '. By contrast, parametric models are often Procustian beds - crude approximations to real world distributions; by abstracting from their detailedness, they often fail to detect informative features such as emerging second modes which might announce a process of income polarisation. Interestingly, in chapter 7 we detect an emerging second mode about the lower income group.

1.6 Chapter 7 : Income dynamics in Germany

The study of income dynamics can be divided into the distinct tasks of examining the shape dynamics of the income distribution, and investigating intra-distributional mobility. The latter aspect is an often neglected dimension of welfare but of particular importance, since it measures the inequality of opportunities, and, according to some commentators, the fairness of the economic system. An unchanging shape of the income distribution may be consistent with polar opposites such as a perfect replication of income positions or their permutation -alternatives amongst which a purely cross-sectional analysis cannot discriminate.

The other point of departure of this chapter is the observation that income mobility in Germany is often regarded to be low. Such a prejudice rests on the assumption that the labour market is segmented, inflexible and immobile; since earnings constitute the main source of income for the majority of the population, incomes should behave similarly. Moreover, the entire structure of the German welfare state is built on the premiss of stability - stable jobs and a consequent low income mobility. But is such conventional wisdom supported?

The shape dynamics are analysed in this paper by two means. The income distributions are estimated directly using (univariate) kernel density methods and inequality indices are computed in order to assess their welfare properties. The analysis of the moving cross-section reveals that income inequality has not (statistically) significantly changed although the sample statistics appeared to suggest an unambiguous increase in inequality as the Lorenz curves shifted out. The density estimates for 1990 and 1991 are unimodal, but in 1992 a second mode emerges in the low income group.

In order to analyse the issue of intra-distributional mobility, two complementary methods are employed. Bivariate kernel density estimates map income transitions at the lowest level of aggregation. But although they constitute a powerful device, they do not permit a rigorous statistical analysis. This is achieved when information is aggregated in transition matrices.

We find that the bivariate density estimates are unimodal. A person's income position is persistent because both contour plots of bivariate kernel density estimates and transition matrices concentrate most mass along the main diagonal. On the other hand 75% of the population experienced income changes in excess of 5% - a fact which is in stark contrast to the popular prejudice of German income immobility. As regards the structure of the transition matrices, they are neither symmetric nor identical as transition probabilities vary with time. This non-stationarity is reaffirmed by computing three popular mobility indices, which suggests that mobility has unambiguously fallen. The useful complementarity between transition matrices and stochastic kernel density estimates are explored, since the latter do not impose arbitrary groupings of information. The contourplots of the latter show the complexity of the income process as different income groups are affected differently; at this level aggregation, there is no universal trend such as a greater concentration about the 45 degree line.

1.7 Chapter 8 : On the non-stationarity of German income mobility (and some observations about poverty)

Chapters 6 and 7 are essentially exploratory in that they discuss the application of various analytical toos for describing income distribution. The next step is to build on these results and to estimate some relevant models in an attempt to explain the observed income dynamics. The point of departure is the conclusion derived in the chapter 7, viz. that income transition probabilities vary with time.

Although the shape of the distribution is remarkably stable in the period 1983 to 1989, mobility has consistently fallen. GSOEP is used in this chapter to test several models which might explain these stylised facts. We proceed in two stages. In the first descriptive and statistical stage we seek a more thorough description of the mobility process. To this end several pure and mixed Markov chains are estimated, and a novel mover-stayer model is proposed which permits time-varying transition probabilities. We show that the process governing income dynamics and intra-distributional mobility are rich and complicated because transition probabilities vary with time and the process exhibits a memory which extends beyond one period.

In the next stage, the strategy is to concentrate on one important income state instead of the entire transition matrix and the chosen income state is poverty. In order to give an economic explanation of the observed transition probabilities, a Markov model with explanatory variables and several duration models (with duration dependent hazards and unobservable heterogeneity) are estimated. The message of the poverty models, be they set in discrete or continuous time, is similar: unemployment is the principal determinant of poverty. Moreover, duration dependence is negative for the long-term poor: the longer the poverty spell, the less likely is the person to escape it. Contrary to results for the US, the household formation process is only of minor importance.

Chapter 2

The evolution of wealth distributions and the economics of information

Abstract:

In this chapter we examine the distributional consequences of informational problems. These are analysed in the context of a dynamic general equilibrium model in which capital markets are modelled explicitly. The economics of information gives rise to powerful mechanisms which lead to persistent inequality. A new economic channel is isolated in this model because standard models, which invoke perfect markets, typically predict the vanishing of inequality. Far from converging to a mass point, the distributional outcome is a nondegenerate distribution.

2.1 Introduction

Economic models based on perfect markets typically predict that initial wealth or income inequalities vanish. However, the assumptions required for such a convergence result are very restrictive. These are relaxed in this chapter as we analyse the distributional consequences of informational problems. Can the economics of information give rise to mechanisms which lead to persistent inequalities ? This question is analysed in the context of a dynamic general equilibrium model in which capital markets are modelled explicitly.

In order to concentrate on this new economic channel, agents are assumed to be identical except for their inherited wealth. They can be productive only if they undertake a project. However, any project is costly and requires the payment of an indivisible admission fee or initial factor outlay. Agents who cannot afford this try to borrow on the capital market. Banks offer a financial contract which specifies an interest rate and a collateral. An incentive problem emerges from the incompleteness of the financial contract: borrowers cannot commit not to undertake a risky activity (which has a lower expected return for banks). But banks can indirectly control the type of project undertaken by appropriate contract design. We analyse the static equilibrium of the model and trace the dynamic consequence of this informational problem.

The plan is as follows: the related literature is reviewed in section 2.1.2 and a discrete and partial equilibrium model due to Stiglitz and Weiss (1987) is discussed in section 2.2 . Section 2.3 presents the general equilibrium model, the mechanism design problem and characterises its static equilibrium. It builds on the revealed preference argument of section 2.2 and demonstrates carefully the dependence of the market equilibrium on the attitude towards risk. Specifically, the case of decreasing absolute risk aversion is examined in section 2.3.1. A continuum of separating contracts emerges in equilibrium. A simulation study is presented in section 2.3.2 which illustrates the previous analysis, and supports various comparative statics exercises. Moreover, a tentative welfare analysis is conducted. Section 2.3.3 briefly considers the case of constant absolute risk aversion. The analysis is rendered dynamic in section 2.4. Section 2.4.1 examines the manner in which the dynamics depend on the model's parameters and initial conditions. A well behaved case emerges for which the dynamics lead to an ergodic distribution. Section 2.5 concludes.

2.1.1 The problem in context

The subject of the dynamic analysis presented here, i.e. a model of dynamically evolving wealth distributions, is not a new one, having originated in the dynamic models of Champernowne (1973). He builds a finite-state discrete and irreducible Markov chain, a model which he later extends to infinite transition matrices, and shows that, given the chosen structure of the transition matrix P, the process converges to a unique stationary distribution irrespective of initial conditions. The latter can be derived analytically as the rescaled eigenvector corresponding to the largest eigenvalue of the stochastic matrix P, viz. 1, and P is chosen to be very well behaved. However, he does not attempt to support his model of distribution with a detailed economic model underpinning the transition equations.

Stiglitz (1969) re-interprets the Solow growth model, similar countries become similar individuals, and shows that in perfect and complete markets the wealth distribution degenerates to a single mass point. However, this exceedingly strong result is qualified by considering disequalising forces, generated by different bequests behaviours, such as primogeniture, or a heterogeneous labour force.

Loury (1981) considers a human capital model and deduces the inequality of incomes from the inequality of abilities. Technically, he abandons a discrete modelling and formulates a model with a continuous state space in which the general theory of Markov operators is used.

The next generation of models is based on the economics of information. Galor and Zeira (1993) obtain a two point limiting distribution by introducing monitoring costs which lead to a difference between borrowing and lending interest rates. Individuals, endowed with insufficient wealth, have to borrow in order to acquire the costly human capital necessary for entering the high earnings sector in a dualistic labour market. Poor agents, however, refrain from borrowing, because the borrowing rates are too high to make this worth their while.

Banerjee and Newman (1991), followed by Aghion and Bolton (1995) and Piketty (1992) also introduce the economics of information into the analysis. Using this framework they motivate capital market imperfections, after noting the extreme assumptions in the previous models of either perfect (capital and insurance) markets or absent markets. This, in conjunction with a non-convexity, is shown to establish a non-degenerate wealth distribution. Inequality of outcome is motivated by either the inequality of opportunities when the non-convexity represents an initial factor outlay, or in the presence of an incentive compatibility constraint. The former paper exploits the classic moral hazard problem of unobservable effort supply, and yields the associated incentive problem for the borrowing entrepreneur: he undersupplies effort because he has to share the marginal returns from effort with lenders. Piketty (1992) uses the same idea but now entrepreneurs sell their projects to insurance firms. In these models perfect insurance is not incentive compatible for it would then be a dominant strategy for the agent to supply no effort -insurance is only partial.

The present chapter is closely related to this new literature on credit market imperfections and on wealth distributions. It extends and generalises the model of Stiglitz and Weiss (1987), summarised below in section 2.1, but employs it for an analysis of evolving wealth distributions. Compared to the papers surveyed above, this chapter enlarges the strategy space of banks and thereby introduces collateral considerations into the analysis. It is believed that collateral plays an important role in financial contracting. Yet, most models assume this issue away. The modelled capital market imperfection is therefore different from the above. The current model is structurally akin to the continuous time model of Banerjee and Newman (1993), because collateral plays an important role on an imperfect capital market with risk neutral agents. Depending on one's initial wealth and the endogenously determined wage rate, agents are subsisting, working, self-employed or employing. Another similarity is that because of this endogeneity, the transitions of the underlying Markov process are not stationary. They show that the long-run equilibrium depends on initial conditions.

A non-convexity must be present in order to establish different investment opportunities, be it a technological one such as an initial factor outlay, or be it one introduced by an incentive compatibility constraint as in Piketty (1992). In this chapter, the former is chosen [1]. Also, the agents' attitudes towards risk receive explicit attention.

[1] As in (Ferreira 1995), we will establish a three-class equilibrium in which agents either are poor, belong to the middle class, or are rich. However, he proposes a partial equilibrium model based on a different mechanism. Two technological non-convexities are present which give rise to two threshold effects, the capital market imperfection arises from a default risk by successful entrepreneurs, borrowers have to put up a collateral but interest rates are fixed exogenously. Both private and public capital enter the production function, and, in the absence of a budget constraint, he shows that an increase in the provision of public

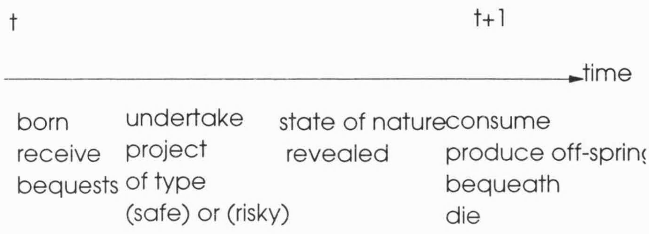


Figure 2.1: The chronology of events

Finally, numerical simulations supplement the analytical results.

2.2 The static partial equilibrium model

Agents, and Projects

The economy is composed of a continuum of risk-averse agents, who are identical except for their inherited wealth. Time is discrete, they live for one period only, and one agent gives rise to one child, so that the population is stationary. The major events in the history of an agent are summarised by Figure 2.1.

Preferences take the form of a standard expected utility function. The cardinal utility function, defined over consumption θ and

capital has an equalising effect on the wealth distribution.

bequests b, takes the form of a concavely transformed Cobb-Douglas function:

$$u(\theta, b) = \varphi\left[\gamma \theta^{\alpha} b^{\beta}\right] \tag{2.1}$$

where $\gamma = \alpha^{-\alpha}\beta^{-\beta}(\alpha+\beta)^{\alpha+\beta}$ is a normalising factor, $\alpha, \beta \in [0;1)$, and φ is a concave transformation function. At present, the transformation φ remains unspecified in order to capture different assumptions about risk aversion.

The indirect utility function is:

$$U(Y) = u(\theta^*(Y), b^*(Y)) = \varphi\left[Y^{\alpha+\beta}\right]$$

as maximising the objective function $u(\theta; b)$ subject to the budget constraint $\theta + b \leq Y$ entails the first order condition $\theta^* = (\alpha/(\alpha+\beta))Y$ and $b^* = (\beta/(\alpha+\beta))Y$. Of course, U is strictly concave with $U' > 0$ and $U'' < 0$. For analytical simplicity let $\alpha + \beta = 1$. The Cobb-Douglas specification gives rise to a mechanistic link between income and bequests as the optimal bequests are linear in realised income Y. Other specifications are conceivable, since, of course, people can live without bequests. However, it is then reasonable to assume that bequests decline in income, worsening the distribution. The Cobb-Douglas form is chosen to facilitate the analysis.

In order to gain access to an investment opportunity, the agent has to pay an indivisible admission fee, or initial factor outlay of c. The agent may then choose whether or not to participate and then

project type	safe	risky	property
success probability	$p^{(safe)}$	$p^{(risky)}$	$p^{(safe)} > p^{(risky)}$
return	$R^{(safe)}$	$R^{(risky)}$	$R^{(safe)} < R^{(risky)}$
			$p^{(safe)}R^{(safe)} > p^{(risky)}R^{(risky)}$

Table 2.1: projects characteristics

between two projects, which differ in terms of their risk and associated pay-off. A failed project pays no return. The safe project $safe$ succeeds with probability $p^{(safe)}$, realising a return $R^{(safe)}$. The risky project is described by $[p^{(risky)}; R^{(risky)}]$, but the expected return on the safe project is higher $p^{(safe)}R^{(safe)} > p^{(risky)}R^{(risky)}$ (although $R^{(risky)} > R^{(safe)}$). An agent, who invests his own funds will therefore always choose the safe project. A borrower, on the other hand, will discount the cost of funds to reflect the probability of bankruptcy. Note that the inequality governing the expected pay-offs is reversed in most models. Table 2.1 collects this information.

Banks, financial contracts, and pay-offs

Banks are competitive, or there is a competitive mutual fund coordinating lending and borrowing. If the funds of an agent -his wealth is denoted by W- are insufficient to gain admission to the investment opportunity he turns to the credit market for finance. Banks have complete knowledge about the type of the loan applicant, so that adverse selection problems are absent.

The financial contract consists of an interest rate r and a collateral requirement $C : (r, C)$. But, as shown below, banks condition r on C, yielding a schedule $r(C)$. For the sake of computational simplicity, attention is confined to standardised contracts, in which borrowers receive a loan of size c, instead of $(c - w)$[2]. In the latter case, the rate of return would also depend on the size of the loan made.

Possibly different contracts are offered. An agent, who cannot meet these terms is denied credit. If the project fails, the agent forfeits his collateral. Consequently the expected return to the bank from a project of type $i \in \{risky, safe\}$ is $\pi = p^{(i)}c(1+r)+(1-p^{(i)})C$. In addition, the agents lends excess funds at the deterministic but endogenously determined lending rate ρ on the credit market because the opportunity costs of lending excess funds is zero.

Let $I \in \{0, 1\}$ be an indicator of the state of nature: '1' if the project is successful, '0' if it failed. The agent then receives the following incomes $Y_I^{(i)}$ from a project of type i.

If he borrows, invests in project of type i, and succeeds, he obtains

$$Y_1^{(i)} = W + R^{(i)} - c(1 + r(W))$$

but if he fails, he forfeits his collateral, and the remaining income $Y_0^{(i)}$ is independent of the chosen project

[2]A similar contract is discussed in Banerjee and Newman (1993) in which "someone with wealth level $w < I$ who wants to become self-employed therefore uses w as collateral and needs to borrow I" (p.281)

$$Y_0^{(i)} = W - C.$$

If the collateral requirement C does not equal the agent's wealth W, the borrower cannot be made fully liable for a failed project - there is some wealth which is inalienable and cannot be collateralised. $W \geq C$ represents thus a limited liability constraint.

If the agent is only a lender

$$Y = W(1 + \rho)$$

Finally, if he is sufficiently rich,$W \geq c$, and would like to undertake the project, he invests in it but lends the excess funds:

$$Y_1^{(i)} = R^{(i)} + (W - c)(1 + \rho) \quad Y_0^{(i)} = (W - c)(1 + \rho)$$

Otherwise his participation constraint is violated and he solely lends. Table 2.2 collects the information about pay-offs.

If the agent undertakes a project of type i, his expected utility is $EU = p^{(i)}U(Y_1^{(i)}) + (1 - p^{(i)})U(Y_0^{(i)})$, defining quasi-concave indifference curves for borrowers in (r, C)-space, because

$$\frac{dr}{dC}|_{EU=const.} = -\frac{1}{c}\frac{1 - p^{(i)}}{p^{(i)}}\frac{U'(Y_0^{(i)})}{U'(Y_1^{(i)})} \tag{2.2}$$

is negative and the sign of d^2r/dC^2 depends on the attitude towards risk. The latter can be seen by taking logarithms of the last term and differentiating, as the sign of the resulting expression depends on the difference of the ratios $U''(Y_I^{(i)})/U'(Y_I^{(i)}), I \in \{0, 1\}$.

type	state of nature	
	success	failure
poor and lend only	$Y_1 = W(1+\rho)$	$Y_0 = W(1+\rho)$
borrow and undertake project i	$Y_1^{(i)} = W + R^{(i)}$ $-c(1+r(W))$	$Y_0 = W - C$
undertake safe project and lend excess funds	$Y_1 = R^{(safe)}$ $-(W-c)(1+\rho)$	$Y_0 = (W-c)(1+\rho)$

Table 2.2: the pay-off structure

A convenient diagrammatic tool for the subsequent analysis of the incentive problem is the switch line introduced in Stiglitz and Weiss(1987), which is an incentive compatibility constraint. It is the locus of contracts which renders the agent with wealth W indifferent between undertaking the safe or the risky project. Above it, he undertakes the risky activity, below it the safe one is undertaken. Consequently, a change in the choice of project takes place. As regards the indifference curve defined by (2.2), this translates into a discrete and reinforcing change of the probability ratio $(1-p^{(i)})/p^{(i)}$ and of the ratio $U'(Y_0^{(i)})/U'(Y_1^{(i)})$. The indifference curve above the switch line is steeper than below it, reflecting the higher failure probability. The utility contours are, therefore, of a winged shape, and not concave. In (r, C)-space, the switch line is upward sloping :

$$\frac{dr}{dC} = \frac{(p^{(safe)} - p^{(risky)})U'(Y_0)}{p^{(safe)}U'(Y_1^{(safe)}) - p^{(risky)}U'(Y_1^{(risky)})} > 0$$

since $p^{(safe)} > p^{(risky)}$ and $U'(Y_1^{(s)}) > U'(Y_1^{(risky)})$. A utility contour

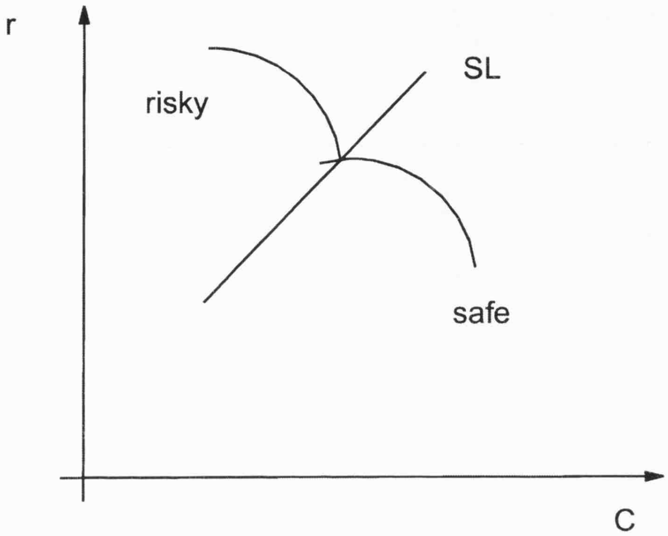


Figure 2.2: The agent's utility contour and the switch line (SL). Above SL, the agent chooses the risky project, below it the agent is better off choosing the safe project.

and the switch line is depicted in Figure 2.2.

2.2.1 The incentive problem in a partial equilibrium model

The model developed so far is rich enough to bring out the main incentive problem, and is discussed in Stiglitz and Weiss (1987). There are just two discrete groups of agents, a middle class and the poor who may offer a collateral of at most C_{poor} and C_{rich}. Stiglitz and Weiss stipulate preferences, which exhibit decreasing absolute risk aversion (DARA), so that for a given contract, richer borrowers would choose riskier projects. The switch line SL of the richer SL(R), then, lies below the switch line of the poor SL(P) (Figure 2.3) whilst the indifference curves of the rich are shallower than those of the poor. (A mean-utility preserving change for the rich, representing a point on the switch line, makes a more risk averse person, i.e. the poorer borrower, worse off. He will then choose the safer project.) Figure 2.3 depicts these properties (an iso-profit line -a locus of equal profits- for banks has been added).

The incentive problem in this model arises from the inability of loan applicants to commit to undertaking a project of a particular type. If faced with a contract above his switch line, banks know that the borrower maximises expected utility by choosing a risky project. But this risky project yields a lower expected return to the bank. As a consequence, banks will offer contracts along the switch line - borrowing rates are sticky. (In equilibrium, a contract cannot lie below the agents' switch line because banks can raise their expected returns by raising the interest rate whilst being certain that the agent still undertakes the safe project.)

Observe also that since $C_{poor} < C_{rich}$ the two groups can easily be separated by demanding a sufficiently large collateral C, which cannot be offered by the poor, $C_{poor} < C \leq C_{rich}$.

Given the conjunction of this incentive problem and the fact that the poor have relatively limited collateral, credit rationing may occur in various forms, depending upon the configuration of preferences

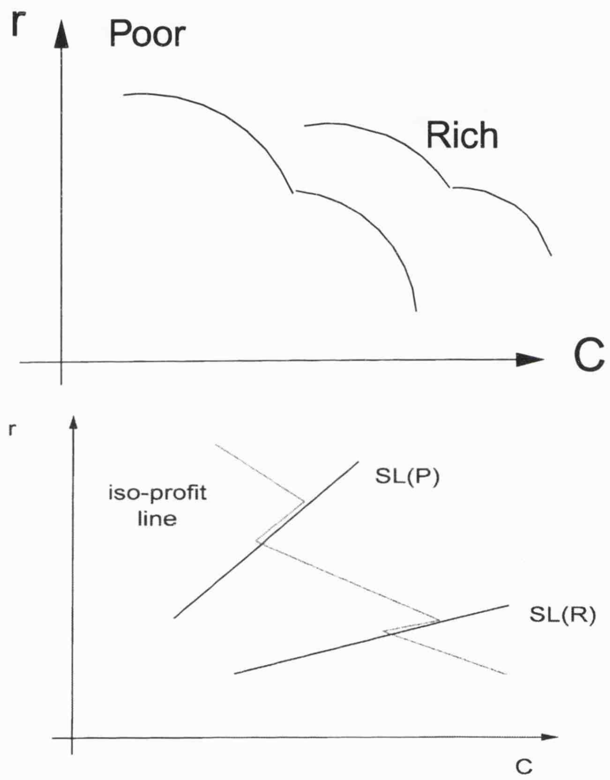


Figure 2.3: With DARA, the indifference curves of the rich are shallower than those of the poor, and the switch line of the rich SL(R) lies below the switch line of the poor.

and the supply of credit. At the maximum feasible collateral (of the poor) there may be a persistent excess demand for credit. The interest rate will not move to eliminate excess demand, for such a Walrasian elimination cannot characterise an equilibrium as the returns to banks would fall. This market shares its generic characteristics with all markets, in which 'quality depends upon price'. As Stiglitz (1987) observes:

"In each of these cases, the story is the same: Because quality (... bankruptcy probability) changes as the price (... interest rate) changes, excess supply or demand may persist without any tendency for price .. to move to correct the market imbalance" [p.7].

Figure 2.4, depicts a situation in which the poor are credit constrained. Although they are the most keen on undertaking a project, they are the most likely to be denied credit. Implicit in this revealed preference argument is the assumptions that it is more profitable to lend to the rich[3] but such a situation is associated with excess supply of credit. The forces of competition drive down the returns along the switch line until all available loanable funds have been allocated. In equilibrium, the two contracts F and E must pay the same expected return to banks, for otherwise funds would be re-allocated, and the two groups consistently undertake the safe activity.

[3]This is an assumption since, in general, collateral C is higher for the rich but the borrowing rate r may be lower.If lending to the rich is not more profitable than lending to the poor, other rationing configurations become possible; cf. Stiglitz and Weiss (1987).

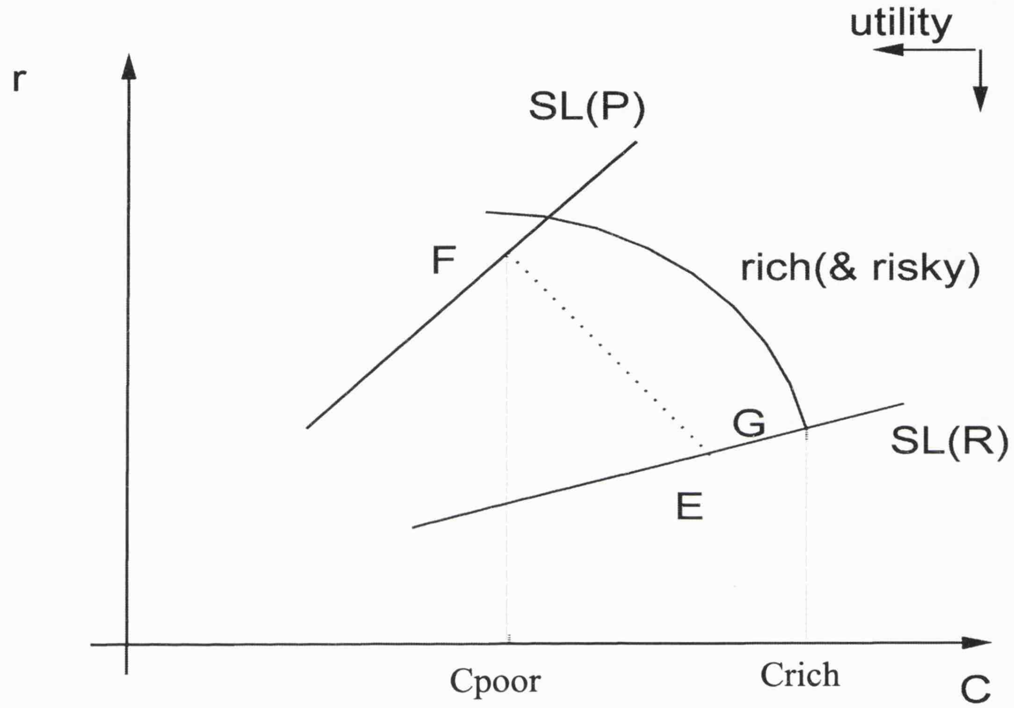


Figure 2.4: Depiction of a separating contract. The poor are offered contract F and the rich obtain E. Both contracts yield the same expected returns to banks. The interest rate will not change to eliminate excess demand.

2.3 The general equilibrium model

Let us build on the insights of the two-class partial equilibrium model to see the implications for a model with a general distribution of wealth. Let $F_t(w)$ denote the proportion of the population at time t with wealth less than w. In the next section we examine the evolution of this distribution, but for the current static analysis the time index has been dropped. For now assume that the initial distribution is, for analytical convenience, well behaved, viz. integrable and piecewise continuous everywhere on its support. Let the support be bounded on $[0, \bar{w}]$. Attitudes towards risk have an important effect on decisions taken by agents, and we analyse two cases. Since this section generalises the Stiglitz and Weiss model, we first give an extensive analysis of the case of decreasing absolute risk aversion considered by them. The case of constant absolute risk aversion is examined next in order to investigate to what extent other forms of risk aversion change the analysis.

2.3.1 Decreasing absolute risk aversion (DARA)

Logarithmic preferences such as $U(c, b) = \ln\{\gamma\theta^{\alpha}b^{1-\alpha} + 1\}$, exhibit DARA because $d\{-U''/U'\}/dY < 0$. Richer borrowers are willing to undertake riskier projects.

The model of the previous section is now generalised by letting the size of the collateral C_i of person i vary with his wealth W_i subject to a limited liability constraint $C_i \leq \delta W_i$, where $0 \leq \delta \leq 1$ is an exogenous parameter which determines the limitation of the defaulting borrower's liability. In fact, let banks demand the maximum collateral $C_i = \delta W_i$, whilst the second instrument -the interest rate- is set to make the contract incentive compatible: undertaking the safe project must give the borrower a greater expected utility than can be gained from undertaking the risky activity $EU_{safe} \geq EU_{risky}$. When met with equality, this constraint yields a solution function $\tilde{r}(W, \eta)$, being the *highest* interest rate for a given collateral δW and model parameters η that a bank can charge, which ensures that the borrower still undertakes the safe project:

$$\begin{aligned} & \left(p^{(safe)} - p^{(risky)}\right) \ln\{(1-\delta)\, W + 1\} \\ = \;& p^{(s)} \ln\{W + R^{(safe)} - c(1+\tilde{r}) + 1\} \\ & - p^{(risky)} \ln\{W + R^{(risky)} - c(1+\tilde{r}) + 1\} \end{aligned} \tag{2.3}$$

As a closed form solution cannot be obtained[4], the constraint will be computed numerically in the simulation study[5]. Despite this,

[4]This problem persists for the entire class of utility functions $U(Y) = (1/(1-a))\, Y^{1-a}$. The logarithmic specification is just its limit when α a tends to 1

[5]Despite this, some analytical discussion is possible. Take the general case of a concave utility function U. To derive the slope of the incentive constraint (2.3), differentiation yields the equation $(\beta_1 - \beta_2)dW = \beta_1 c d\tilde{r}$ where $\beta_1 = p^{(safe)}U'(Y_1^{(safe)}) - p^{(risky)}U'(Y_1^{(risky)}) \geq 0$ and $\beta_2 = (p^{(safe)} - p^{(risky)})(1-\delta)U'(Y_0) \geq 0$. As the liability of the borrower becomes greater, banks extract a larger collateral for a given wealth level, and in the limit $\delta \uparrow 1$ and $\beta_2 \downarrow 0$.

the effects of various parametric changes can be examined. Increasing the return on the safe project, for instance, it can be shown by differentiating (2.3) totally with respect to $\tilde{r}$ and $R^{(safe)}$ that $d\tilde{r}/dR^{(safe)} > 1$.

What is the set of contracts offered in equilibrium ? Let ω denote the cut-off point on the wealth line below which agents are either credit constrained or optimally choose not to borrow (see discussion below on the determination of ω). If the incentive constraint defined by (2.3) yields an upward sloping schedule in $(r; W)$-space, then it is always more profitable to lend to the richer loan applicants rather than to the poorer since not only is his automatically fixed collateral level higher, but also the incentive compatible interest rate the bank can extract from him.

In equilibrium, every offered contract must fulfil two requirements: all contracts must yield the same expected return, say Π_ω, the lowest return on a contract associated with wealth level , the cut-off point ω. The equal profit condition is:

$$p^{(safe)}c(1+\tilde{r}(\omega)) + (1-p^{(safe)})\delta\omega = p^{(safe)}c(1+r) + (1-p^{(safe)})\delta W \tag{2.4}$$

for all W such that $\omega < W < 1$. This equation defines a solution function $r(W, \eta)$ where the vector η collects all the parameters. The con-

In this case $d\tilde{r}/dW > 0$. However, as the liability of the borrower becomes increasingly more limited, as δ decreases the slope of the incentive constraint can switch signs.

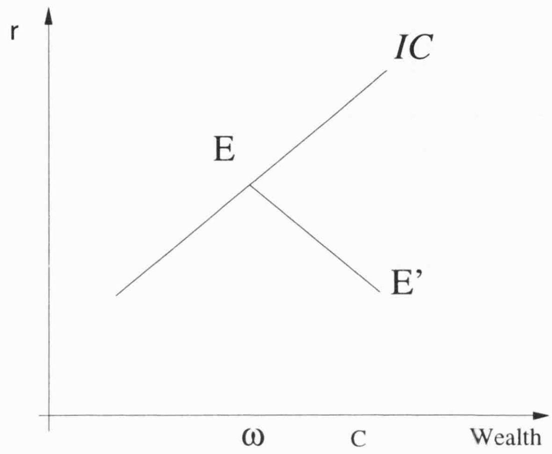


Figure 2.5: The set of incentive compatible contracts lies on the segment EE'. The line IC depicts the solution to the incentive constraint defined by equation (2.3).

tract offered to person i whose wealth is W_i is thus $(r_i(W;\eta);\delta W_i)$. As the second equilibrium requirement, the offered contract must also satisfy the person's specific incentive compatibility constraint. But the set of contracts defined by equation (2.4) is indeed incentive compatible since the interest rate has been reduced, thus increasing the attractiveness of the safe project and relaxing the incentive constraint. In consequence, the set of contracts offered in equilibrium yield expected profits Π_ω, and lie on a downward sloping schedule defined by equation (2.4), illustrated in Figure 2.5 as line EE'.

Having thus characterised the set of actual contracts offered, it is possible to examine the behaviour of the participation constraint for a potential loan applicant. The borrower optimally undertakes the project when its associated expected utility exceeds the expected utility derived from lending at ρ. From this we obtain the participation constraint:

$$\begin{aligned} & p^{(safe)} \ln\{W + R^{(safe)} - c(1 + r(W, \eta)) + 1\} + \\ & (1 - p^{(safe)}) \ln\{(1 - \delta)\, W + 1\} \\ \geq \; & \ln\{W(1 + \rho) + 1\} \end{aligned} \qquad (2.5)$$

This inequality yields an indicator $j_W \in \{0, 1\}$ depending on the wealth level, which states whether or not the agent would like to undertake the project given the financial contract and thus apply for a loan. How does equation (2.5) behave ? Given the cut-off level ω, as wealth increases, so does the associated collateral and the required interest rate falls. Thus the left member of (2.5) increases. But, given ω, ρ remains unchanged and so the right member increases as well. The relative effect depends on the parameters and the initial distribution and is examined in the simulation study of section 3.2. Note, however, that economic growth leads to a reduction of the lending rate ρ as loanable funds become less scarce, reducing the participation constraint for any given wealth level. If the participation constraint for the loan applicant is satisfied, then so is the participation constraint for the rich $W \geq c$ because they do not have to spend resources on repaying a loan. In general, meeting such a type of participation constraint requires that agents are not too risk averse, since they face a (low) chance of bankruptcy. The pay-off from the successful project must exceed their risk premium.

Given these properties of the model discussed so far, it is possible to complete the description of the financial market and to classify agents according to their wealth. The demand for loans comes from the mass of agents, who inherit a wealth level less than the admission fee for investment. However, since the supply of loanable funds may prove insufficient, and agents with a higher collateral are in front of poorer agents in the credit queue, the poorer agents may be denied credit. All they can do is lend at the prevailing depositors' rate ρ. Let Ω denote the point at which all loanable funds are exhausted. The supply of funds comes from the poor and from the rich.

This behaviour implies the following quantity constraint on the financial market since funds are firstly allocated to the richer loan applicants

$$\int_{\Omega}^{c} cdF(w) \leq \int_{0}^{\Omega} wdF(w) + \int_{c}^{\infty} (w-c)dF(w) \qquad (2.6)$$

with $0 \leq \Omega \leq c$. Ω is a measure of the credit capacity of the economy, for the richer the economy, the lower is Ω. If the economy is initially too poor, $\Omega \geq c$, the financial market does not exist, and only the rich undertake the safe project (In this case the economy will have to grow before a sufficient surplus of funds is accumulated). Whether ω=Ω depends on the participation constraint. If ρ is not too high, then the person with wealth level Ω would like to undertake the

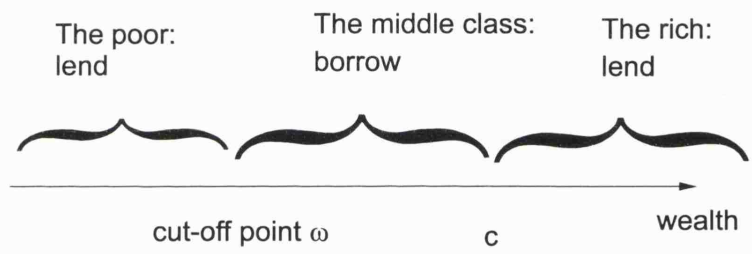


Figure 2.6: The type of agents ordered on the wealth line

project and thus $\omega = \Omega$. However, if the participation constraint is violated, then $\omega > \Omega$.

Finally, the lending rate is determined endogenously from the zero-profit condition for banks. Equating expected costs and expected returns, making use of the law of large numbers (which requires that the economy is sufficiently large), and given the cut-off point ω, yields ρ as the rate of return on lending one unit:

$$\begin{aligned} &\rho\left[\int_0^{\omega} w dF(w) + \int_c^{\infty}(w-c)dF(w)\right] \\ =& \int_{\omega}^{c}\left[p^{(safe)}c(1+r(w)) + (1-p^{(safe)})\delta w\right] dF(w) \end{aligned} \tag{2.7}$$

with $\rho \geq 0$. Note that when loan supply increases, ceteris paribus, ρ falls, so that the equation is well-behaved.

Summary of the model

The model has a difficult structure because of the many simultaneities, and non-linearities. The incentive compatibility and participation constraints introduce important inequalities into the analysis.

The nature of the financial contract arising from the incentive problem sorts agents into classes: the poor lenders, middle class borrowers and entrepreneurs, and the rich. (2.6) determines the credit capacity Ω of the economy. The global incentive compatibility constraint (2.3) constrains the set of equilibrium contracts. If it is upward sloping, then it is always more profitable to lend to richer loan applicants and a credit queue emerges. The least profitable contract is associated with the wealth level ω and must lie on (2.3), viz. $\{\tilde{r}(\omega); \delta\omega\}$. All other contracts are determined by the equal profit condition (2.7). This contract curve is downward sloping. Given this set of contracts, (2.6), and the cut-off level ω, the lending rate ρ is determined by (2.7). Finally the participation constraint gives a critical level $\tilde{\omega}$ above which agents would like to undertake the project and thus borrow. ω is determined by $\max\{\Omega, \tilde{\omega}\}$.

2.3.2 The simulation and some comparative statics

Despite the complex structure of the model, its simulation is not too difficult since the underlying implicit functions determining Ω in equation (2.6) and $\tilde{r}(w, \eta)$ in equation (2.3) are monotonic, and their values can be calculated using binary search algorithms. Of principal importance is the verification of the participation constraint of loan applicants, the behaviour of which at times may appear surprising as the discussion below demonstrates. The computer simulation offers a convenient method by means of which to examine how the properties of the model depend upon: the initial parameters, the initial distribution, the degree of risk aversion, and the cost of the project. Finally, a tentative welfare analysis is conducted.

The role of the initial distribution

The properties of the participation constraint (2.5), it turns out, are of fundamental importance, affording characteristics according to which economies may be classified. First, I examine the extent to which the static equilibrium depends on the initial wealth distribution and the resulting lending rates. This analysis gives rise to two types of economies.

Type I: The economy may be endowed with insufficient wealth. In this case prices are too high and no trading takes place. Inaction is the only equilibrium: the financial market is non-existent, only the rich undertake the safe project, and everyone else is condemned to inaction.

Type II: A more moderate effect may emerge in richer economies, but it is a manifestation of the same driving force. If the lending rate is sufficiently high (but not as high as in the previous case), then the attractiveness of undertaking the project over mere lending decreases as the wealth of a loan applicant increases. The participation constraint (2.5) becomes less relaxed, even though the required interest rate $r(w)$ falls as wealth increases. This situation implies a form of rationing: although the poorest agents could benefit most from undertaking the project and are therefore the most keen on obtaining credit, they are also the most likely to be denied credit.

An example of an economy of this type is given by the parameters $c = 1, \delta = 0.8, (p^{(safe)}; R^{(safe)}) = (0.9; 1.9), (p^{(risky)}; R^{(risky)}) =$

(0.5; 2.8), and an initial distribution which is discrete and uniform ranging from 0.01 to 3 on a wealth grid with $\Delta w = 0.01$. In this case, all persons whose wealth lies below the admission fee $c = 1$ could be allocated credit, Ω =0.01. All participation constraints are satisfied, $\omega = \Omega$, but in the above manner, so that (2.5) becomes less relaxed in wealth. Figure 2.7.A depicts this situation. The lending rate is ρ=0.527, whilst the highest borrowing rate is $r(\omega)$= 0.207. Figure 2.7.A depicts the set of equilibrium contracts for this economy.

A change in risk aversion

A change in risk aversion can be modelled by changing the indirect utility function to $V = \ln\{Y + \epsilon\}$, where $\epsilon > 0$ is a constant, so that risk aversion decreases as ϵ increases because $d(-V''/V')/d\epsilon < 0$. How does this affect the incentive compatibility constraint (2.3) ? The intuition suggesting that the schedule moves down is the same which underlies Figure 2.3 [6]. This is borne out by the simulations. As a consequence, the lending rate falls for a given wealth distribution. For instance, using the above configuration for the type II economy (with no binding credit constraints), changes in the risk aversion parameter ϵ yield the results collected in Table 2.3.

[6] Analytically $d\{d\tilde{r}/dw\}/d\epsilon$ is ambiguous, since differentiating the expression $d\tilde{r}/dw$ derived in the previous footnote yields a messy expression, the first term of which is negative, the second of which is positive. However, as $\delta \uparrow 1$, $d\{d\tilde{r}/dw\}/d\epsilon$ becomes unambiguously negative.

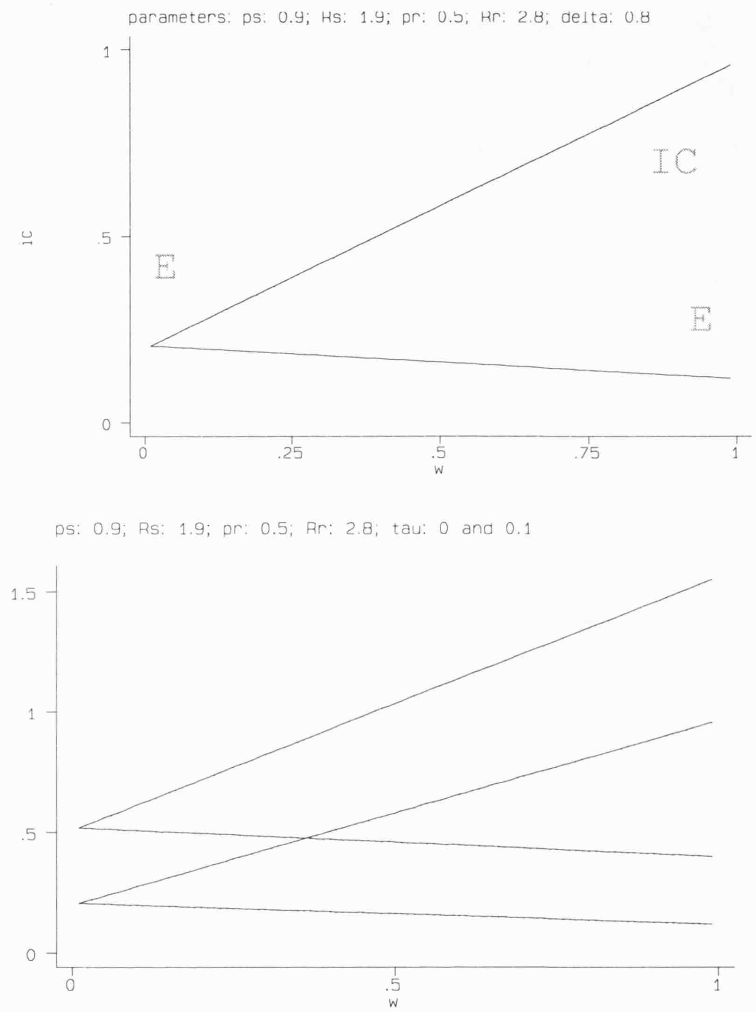


Figure 2.7: Some simulations.

The upper diagram depicts Figure 2.5 for the chosen parameters. On the IC line the incentive constraint (2.3) is met with equality. The equilibrium contracts lie on the line EE. For the chosen parameters and the initial distribution, there is no credit rationing.

Lower diagram: government intervention leads to higher interest rates. The IC schedule for the subsidised economy lies above the one for the unsubsidised economy.

ϵ	0.9	1	1.1
ρ	0.536	0.527	0.5126

Table 2.3: changes in the risk aversion parameter

Changing the cost of the project

As the cost of the project c increases, the credit capacity of the economy Ω cannot increase since funds become scarcer. What happens to the incentive constraint (2.3)? As before, the behaviour of $d\{d\tilde{r}/dw\}/dc$ is somewhat ambiguous, being composed of two opposing effects. However, as $\delta \uparrow 1$, it becomes unambiguously negative, so the schedule swivels down. This should follow, since an increase in the project cost, be it a direct increase in the cost or an indirect one through the rise in the interest rate, makes the risky project more attractive. (Pick a contract that satisfies the incentive constraint with equality. Holding the interest rate constant, an increase in c induces undertaking the risky project. Thus the interest rate has to be reduced in order to induce undertaking the safe project.)

The simulation reflects these consequences. For the economy of the previous example c has increased from 0.8 to 1.2, and the results are collected in the Table 2.4. Note, however, that the economy is sufficiently rich, so that it can absorb this increased project cost without reducing its credit capacity. $\tilde{r}(\omega)$ falls but ρ increases. Thus the direct effect of an increase in c outweighs the decline in the in-

c	0.8	0.9	1.0	1.1	1.2
$\tilde{r}(\omega)$	0.508	0.34	0.207	0.097	0.006
ρ	0.424	0.475	0.527	0.528	0.532

Table 2.4: changes of the project cost c

terest rate charged, so that there is an overall increase in the lending rate.

Welfare analysis

Can a government subsidy, which reduces the cost of investment, increase social welfare? Any policy objective has to be evaluated in the light of the constraints on the system: the subsidy has to be financed through taxation, and a balanced budget needs to be maintained. The analysis of such a situation depends on a host of assumptions, in particular on the assumed tax schedule, the social welfare function employed for the evaluation of the redistributive consequences, and on the initial conditions of the economic system. The subsequent discussion reflects this level of generality, and only aims at illustrating some points.

Let all potential borrowers, i.e. all persons with wealth level $w < c = 1$ be subsidised by an amount Δ, whilst a lump-sum tax τ is levied on all successful projects. The balanced budget constraint requires $\Delta = \{p^{(safe)} \int_{\Omega}^{\infty} dF / \int_{\Omega}^{1} dF\}\tau$ and thus depends on the wealth distribution. Such a policy is redistributive from the rich to the poor.

Pareto improvements are not possible in this static model since not everyone can feasibly experience a net gain in utility. How does this policy affect the incentive problem (2.3) ? Differentiating this schedule with respect to Δ results, as before, in ambiguous results, which disappear as $\delta \uparrow 1$. In the latter case $d\{d\tilde{r}/dw\}/d\Delta > 0$, being a similar effect as caused by a reduction in the project cost because this redistributive policy reduces the net cost of the project for borrowers.

The simulation exercise confirms these observations. Continuing the above example with c=1, levying a lump sum tax of τ=0.1 leads to a reduction of the project cost by Δ=0.27. However, this only aggravates the incentive problem, a situation which is depicted in Figure 2.7.B. The IC schedule for the subsidised case lies above the schedule for the unsubsidised economy, and thus do all equilibrium contracts. Assuming that no person is credit constrained, this rise in the borrowing rate reduces welfare. Only in the presence of credit constraints can there be net welfare improvements. If previously credit constrained persons may now undertake the project, then there utility gain may outweigh the utility loss from an increase in the borrowing rates and the tax levy. But this scope for aggregate utility gains depends crucially on the concavity of the social welfare function. The less concave it is, the less scope for welfare improvements there is.

2.3.3 Constant absolute risk aversion (CARA)

If the concave transformation function φ in (2.1) is chosen to be exponential, then preferences are $u(\theta, b) = -\exp\{-\gamma\theta^{\alpha}b^{\beta}\}$, and exhibit CARA for, when $\alpha + \beta = 1$, $-U''/U' = 1$. Let the cost of the project be $c=1$ for notational simplicity. The principal change to the above analysis is that the incentive constraint can now be calculated explicitly. When this constraint is met with equality $EU_{(risky)} = EU_{(s)}$, which implies

$$\tilde{r} = \ln\{\frac{(p^{(safe)} - p^{(risky)})\exp\{-(1-\delta)\,w\}}{p^{(safe)}\exp\{-(w + R^{(safe)} - 1)\} - p^{(risky)}\exp\{-(w + R^{(risky)} - 1)\}}\}$$

which can be simplified to

$$\tilde{r} = \ln\tilde{\alpha} + \delta w - 1 \quad \text{where } \tilde{\alpha} = \frac{p^{(safe)} - p^{(risky)}}{p^{(safe)}\exp\{-R^{(safe)}\} - p^{(risky)}\exp\{-R^{(risky)}\}}$$

This schedule is upward sloping in $(r; w)$-space. Hence the conclusions derived in the previous analysis remain fully intact.

2.4 Dynamically evolving wealth distributions

Let F_t be the distribution of wealth at time t. It will be assumed, at first, that F_t is bounded on $[0, \hat{w}]$ for all t, but an argument presented below demonstrates that this boundedness is a property of the model, arising from a convergent difference equation in wealth.

Let $(R, \mathfrak{R}, \mathfrak{p})$ be a probability space, where $\mathfrak{R}$ is the σ-field of Borel subsets of the real line. Agents bequeath a fraction $(1-\alpha)$ of their realised income to their offspring. Consequently, the evolution of the wealth distribution $F_{t+1} = T^*(F_t)$ is determined recursively by the transition equations for dynasties, viz. their bequests. These define a Markov process since only the preceding wealth position is relevant for the determination of the current one. The stochastic process is perhaps best visualised as a generalisation of a finite state discrete irreducible Markov chain with transition matrix P. In such a case the stationary distribution π is derived as the fixed point of the equation $\pi = \pi P$, or the rescaled eigenvector corresponding to the largest eigenvalue of the stochastic matrix P, viz. 1. For a model with a continuous state space, the technicalities become more complex and abstract theory of operators needs to be employed [7].

The Markov operator is defined as a convolution

$$(Tf)(z) = \int f(z')Q(z, dz') \quad all \quad z \in R$$

where $Q(a, A) = \Pr\{z_{t+1} \in A | z_t = a\}$ is the transition function. $(Tf)(z)$ is the expected value of the function f next period conditional on the current state z. The adjoint of the Markov operator, denoted by T^*, is defined by means of the inner product relationship $< Tf, \lambda > = < f, T^*\lambda >$. Expanding the inner product yields

[7]See Chung (1960) for an extensive but concise treatment of Markov chains. For a good general introduction to operator theory in a Hilbert space setting see Young (1988). Futia (1982) surveys opeator-theoretic techniques and limiting theorems. Lucas and Stokey (1989) collect convergence results and other useful theorems. The classic but voluminous treatise is Dunford and Schwartz (1957).

$$(T^*\lambda)(A) = \int Q(z, A)\lambda(dz) \quad all \quad A \in \mathfrak{R}$$

It is the probability measure over the state next period given that λ is the probability measure over the current state. Yet, since the model is a general equilibrium model, the lending rate ρ is determined endogenously, and enters the pay-off and thus the (linear) transition equations. As a consequence, the Markov process is initially not stationary, and the operator T^* is time dependent. Thus the evolution of the wealth distribution is recursively determined as $F_{t+1} = T_t^*(F_t)$.

However, if the economy is well-behaved and grows (as discussed below), the lending rate ρ falls over time, rendering the transitions stationary. This convergence can be established by an argument by contradiction. If ρ does not converge to $\underline{\rho}$, then the wealth of the rich grows without bound. But if this happens, then, on the other hand from (2.7), $\rho \downarrow \underline{\rho}$, contradicting the previous assumption. Moreover, as the economy grows, the loan capacity increases, reducing any previous (if any) credit constraint. Finally, if the economy is stationary, then the transitions remain stationary in the succeeding periods [8].

[8]For a rigorous proof of this proposition in a similar context see Aghion and Bolton (1995).

2.4.1 Parameter values and the nature of the dynamics

The next paragraphs examine the manner in which the model's parameters and initial conditions determine the nature of wealth dynamics. Note that because of the special specification of the utility function, wealth transitions are linear in pay-offs as parents bequeath a fraction $(1-\alpha)$ to their offspring. For simplicity let the cost of the project be $c = 1$.

<u>The rich and wealth growth</u>

Since the rich lend excess funds at the endogenously determined lending rate ρ, their transition functions are not stationary. For what parameter values will there be (initial) wealth growth? Their wealth grows if

$$E\{w_{t+1}|w_t > 1\} - w_t = (1-\alpha)\left(p^{(safe)}R^{(safe)} + (1+\rho)(w_t - 1)\right) - w_t > 0.$$

In the worst case, if $(1+\rho) > (1-\alpha)^{-1}$ then we require $p^{(safe)}R^{(safe)} > (1+\rho) > (1-\alpha)^{-1}$ for there to be growth. In consequence, the returns to the project must be high, whilst the initial distribution must not be too rich, so that initially the lending rate ρ is high as well. For instance, if parents bequeath a fraction 0.6 of their wealth, α=0.4, and the following is required: $\rho > 2/3$ and $p^{(safe)}R^{(safe)} > 5/3$. If $p^{(safe)} = 0.9$, then $R^{(safe)} > 1.852$ must hold.

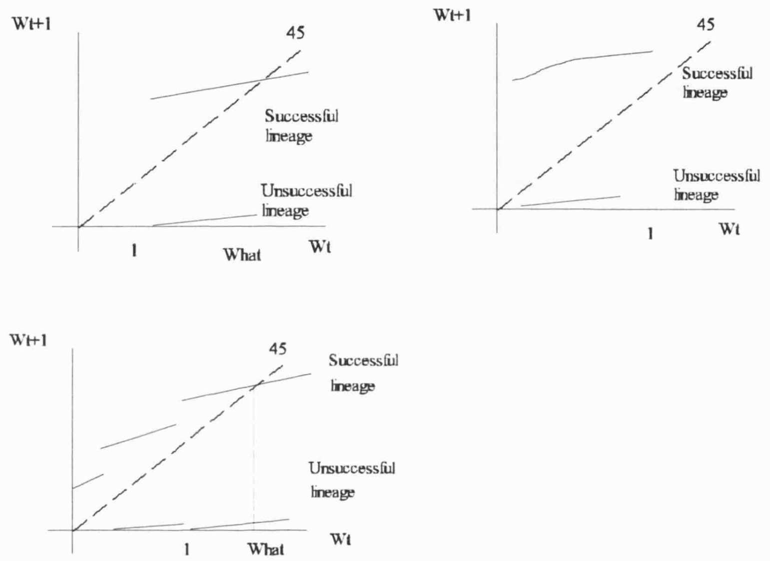


Figure 2.8: Upper diagram: Recursion diagram for the rich

Right diagram: Recursion diagram for the middle class

Lower diagram: Recursion diagram for the well-behaved case

What type of recursion diagram emerges for the rich ? Assume ρ has fallen to zero, and concentrate on the lineage which is always successful (the probability of which event falls to zero $p^{(safe)} \downarrow 0$). If $(1-\alpha)\left[R^{(safe)} + (w_t - 1)\right] - w_t > 0$, then the transition function for this group lies initially above the 45 degree line. This condition requires $R^{(safe)} > \alpha(1-\alpha)^{-1}w_t + 1$. In particular at $w = 1, R^{(safe)}\alpha(1-\alpha)^{-1} + 1$. If $\alpha = 0.4$ then $p^{(safe)}R^{(safe)} > 5/3$ must hold. Note, however, that the above condition cannot hold for all wealth levels w, since it defines a convergent difference equation. Thus wealth above a certain point, say $\hat{w}$, cannot be sustained. This argument justifies restricting the support of the wealth distribution F. The recursion diagram is depicted in Figure 2.8.A.

The middle class

Since the incentive constraint (2.3) is time invariant, if the economy is sufficiently rich, $\omega = 0$, the transition function for members of the middle class is stationary. How does the recursion diagram look ? Focusing again on the successful lineage, its transition function lies above the 45 degree line if $(1-\alpha)[w_t + R^{(safe)} - (1 + r(w_t; R^{(safe)}))] - w_t > 0$. Since the incentive constraint (2.3) defines an upward sloping schedule in (r, W)-space, this constraint is at its severest at $w = 1$. Whence we required that $R^{(safe)} > \alpha(1-\alpha)^{-1} + (1 + r(1; R^{(safe)}))$. Continuing the above example, the requirement is $R^{(safe)} > 5/3 + r(1; R^{(safe)})$ [9]. The recursion diagram for this case is depicted in Figure 2.8.B.

However, if the above parameter restriction is not met, the dynamics are not "well-behaved", the ergodic theorem is inapplicable, and the limiting distribution depends on initial conditions. In terms of the recursion diagram, the transition function for the successful lineage then crosses the 45 degree line.

The poor

The wealth transitions of the poor, as of the rich, are not stationary because ρ is endogenously determined. However, if the latter has fallen sufficiently, they decumulate their assets until, eventually, they occupy a mass point at $w = 0$. Since the modelling for this case is unrealistic, a more realistic model incorporates a subsistence activity. Introducing such a subsistence activity changes the participation constraint of the middle class borrowers, if undertaking the project or the subsistence activity are mutually exclusive actions. In this case the expected utility derived from undertaking the project must exceed the utility derived from the subsistence activity and lending. If the return on the subsistence activity s is sufficiently high (in the worst case) $(1-\alpha)[w_t + s] - w_t > 0$, the recursion diagram is well behaved: the transition function then lies above the 45 degree line.

[9] Note, however, a subtlety in this condition since as $R^{(safe)}$ increases so does the contracted borrowing rate.

The well-behaved case and the ergodic distribution

If the above parameter restrictions are met, the resulting recursion diagram is depicted in Figure 2.8.C. An agent attains any wealth level within the support of the distribution with positive probability and in finite time. Moreover, the transition functions and thus the Markov operator become stationary. Thus, in this well behaved case the ergodic theorem (see appendix) applies and a unique invariant distribution emerges.

2.5 Conclusions

The chapter has analysed the dynamic consequences of a non-commitment problem in the credit market and has shown that the behaviour of financial intermediaries brings about a new economic channel through which inequalities are made persistent. The inequality of opportunities leads to persistent wealth inequality, the dynamics of which - a Markov process - depend upon parameters and initial conditions. (although one well behaved case emerges for which the ergodic theorem is applicable).

Models with perfect markets typically predict that initial inequalities vanish. For instance, Stiglitz (1969) considers "a simple model of accumulation, with a linear savings function, a constant reproduction rate, homogeneous labour, and equal division among one's heirs.

In such an economy, if the balanced growth path is stable, all wealth and income is asymptotically evenly distributed." (p.382). But far from converging to a point mass, the outcome here is a nondegenerate distribution. Thus the economics of information has an important contribution to make to distributional analysis; working through the channel of financial intermediaries, this emphasises again the importance of analysing the role of intermediate institutions for the final distributional outcome -the general theme of this dissertation.

A quite different institution is examined in the next chapter. The theory of Markov processes has been employed here in the context of an abstract model. Chapters 7 and 8 examine the applicability of the former in an empirical and statistical context.

2.6 Appendix: the ergodic theorem and the well-behaved case

The ergodic theorem applies in the well-behaved case, since agent attains any wealth level within the support of the distribution with positive probability and in finite time. See Aghion and Bolton (1995) for a detailed and explicit construction of the Markov operator. Consequently, the conditions for the following theorem are readily verified, establishing the existence of a unique invariant distribution. One form of the convergence theorem is given in Lucas and Stokey (1989) (theorem 11.12 , p.350), which builds on a particular structure of the transition equations:

Condition M (p.348): There exists $\varepsilon > 0$ and integer $N \geq 1$, such that, for any $A \in \mathfrak{R}$, either $P^N(s, A) \geq \varepsilon$, or $P^N(s, A^c) \geq \varepsilon$, $\forall s \in R$.

Theorem (p.350): Let $(R, \mathfrak{R})$ be a measurable space; let $\Lambda(R, \mathfrak{R})$ be the space of probability measures on $(R, \mathfrak{R})$, with the total variation norm [10]; let P be a transition function on $(R, \mathfrak{R})$; let T^* be the adjoint of the operator associated with P. If P satisfies condition M for $N \geq 1$ and $\varepsilon > 0$, then there exist a unique probability measure $\lambda^* \in (R, \mathfrak{R})$ such that:

[10] Recall the total variation norm defined as $||\lambda|| = \sup \sum ||\lambda(A_i)||$ where the sup is taken over all finite partitions of the state space into disjoint subsets. It is the norm of the Banach dual, since T^* maps the Banach space of bounded finitely additive functions defined on the space into itself.

(*) $||T^{*Nk}\lambda_0 - \lambda^*|| \leq (1-\varepsilon)\,||\lambda_0 - \lambda^*||$ all $\lambda_0 \in (R, \mathfrak{R})$, k=1,2,..

Conversely, if (*) holds, then Condition M is satisfied for some $N \geq 1$ and $\varepsilon > 0$.

Chapter 3

On the performance of social benefit systems

Abstract: The chapter analyses the performance of unemployment benefit systems in a search theoretic framework. The criteria of evaluation comprise the alleviation of poverty and the reduction in income inequality. Diversity of opinions about these criteria is explicitly allowed for. Also, the trade-off between the attainment of social objectives and work incentives is examined.

3.1 Introduction

Social insurance is a prominent institution in developed economies, designed to protect economic agents against income risk. The form of social insurance schemes found in practice varies considerably: unemployment benefit in the UK is a flat rate, whereas it is earnings related on the Continent and the US; the replacement rates and the methods of finance differ. This chapter analyses the performance of alternative unemployment benefit systems in a search-theoretic framework. The relative performance of flat-rate and earnings-related unemployment benefits will be assessed in the light of policy objectives such as the reduction in inequality and the alleviation of poverty. However, these policy objectives may not command a universal consensus because of either a diversity of opinion or an intrinsic arbitrariness in the parameters characterising the social welfare function [1]. So we might ask whether the ranking of the benefit regimes depends on the parameters of the social objective; although people may disagree about parameters, could they agree on a ranking ?

There is also a potential trade-off between the equity objective of poverty alleviation and the efficiency consideration of work incentives. A greater benefit might increase the income of an unemployed

[1]As Atkinson (1993a) observes in the context of poverty alleviation: "Such a 'sharp' representation of the social objective may not, however, be universally accepted. There may well be disagreement about the location of the poverty line ... Alternatively, there may be agreement about the location of ..[the poverty line], but concern for the non-poor, or the group close to the poverty line"(p.17).

beneficiary and thereby reduce the difference between his income and the poverty line. For a person who remains in poverty, this increase reduces his poverty. But an increase in benefits reduces the incentives to work (particularly for those persons with a current low job productivity). Consequently, unemployment might rise, possibly increasing the numbers of the poor, raising aggregate poverty. This potential trade-off is examined in a general equilibrium setting. Again, does the resolution of this trade-off depend on the parameters of the social objective, or is it unambiguous for all admissible parameters ?

The chapter is structured as follows: Section 3.1.2 briefly reviews the related literature. The model is presented in section 3.2. It juxtaposes a flat-rate and an earnings-related unemployment benefit regime in a general equilibrium framework. Incentive problems of the benefit regimes are also examined. Section 3.3 assesses the relative performance of the benefit regimes, and analyses the conditions under which an initial ranking of the benefit regimes is reversed. In section 3.3.1 the evaluation criterion is poverty, and in section 3.3.2 it is inequality. Within the (limited) context of the current search-theoretic framework, it is shown that no benefit regime dominates its competitor in all circumstances. Section 3.3.3 examines whether there is an equity-efficiency trade-off. As this section makes clear, preserving incentive compatibility for its own sake is a value judgement, and its normative character ought to be made explicit. Section 3.4 concludes.

3.1.1 The context of the problem

The nature of the risk analysed in this chapter is that of a temporary job loss, engendering the institutional response of unemployment insurance. Pension schemes attempting to counter the risk of a permanent job loss through age or disability have been examined elsewhere (e.g. Diamond and Mirrless (1986)). Poynter and Martin (1995), Habib (1995) , and Schluter (1995) provide extensive examinations of the complexities governing the British, the French, and the German social insurance system.

These institutional features have been relatively neglected in the recent literature on social insurance schemes. Some chapters completely ignore the institutional rules, whilst taking them explicitly into account dramatically reverses the implications of some popular models (Atkinson (1990)). An example is the treatment of eligibility conditions in the Shapiro and Stiglitz (1984) efficiency wage model. Since, in practice, shirking automatically disqualifies the claimant from benefit entitlement for a non-trivial period, their shirking condition is simply not applicable. Atkinson and Micklewright (1991) develop this criticism of neglecting institutional considerations in economic modelling. The incentive problems engendered by the benefit system are further analysed in Besley (1990), who compares the relative performance of a means-tested benefit, which tops up incomes in order to reach a pre-defined poverty line, with a universal benefit, paid to all persons in the economy, even to the richest person. Some authors, like Besley and Coate (1992), model benefit institutions in a historical fashion. They design a revelation mechanism by subjecting the benefit claimant to a sufficiently large work requirement, the result of which bears a strong resemblance to the Poor Laws in Britain and practices in colonial India (Dreze (1990)).

Easley, Kiefer, and Possen (1985) use a two-state, two-period general equilibrium model with two heterogeneous risk averse agents in order to analyse the (relative and joint) performance of unemployment insurance and negative income tax systems. Their numerical simulations suggests that under certain parameter configurations and functional forms, both programmes bring about Pareto improvements. Other researchers have incorporated efficiency wages into general equilibrium models of unemployment: when workers' effort depends on the relative remuneration of capital and labour, Agell and Lundberg (1992) show that any tax policy which increases the wage-rental rate leads to a reduction in unemployment.

The model developed in this chapter is based on a standard search-theoretic framework as described in Pissarides (1990). Its principal attraction stems from the fact that it endogenises the risk of losing one's income. This is achieved by modelling trade in the labour market as uncoordinated, time consuming and costly for both workers and firms. A congestion or thin market externality will be present in most equilibrium conditions, their levels depending on the number of workers and firms engaged in search. With the unemployment rate thus endogenised, its level will reflect the incentive structure of the benefit system. It is then possible to examine to which extent unemployment is caused by the incentive structure under operation, rather than by the stochastic nature of the exogenous shocks. Involuntary unemployment can be distinguished from 'voluntary' unemployment. It has been observed that "(t)he optimum taxation models developed to date are not satisfactory in this regard, since the treatment of the labour market is insufficiently developed" (Atkinson (1989), p.42). The model employed in this chapter is an attempt at this in the context of social insurance. In using this framework, the present analysis is akin to Atkinson (1990)'s, in that a search-theoretic framework is also employed. However, he analyses these issues within a model of a segmented labour market: a primary sector job offers high wages and unemployment insurance, whereas the low paid jobs in the secondary sector are uninsured.

3.2 The model

This section spells out a standard search-theoretic framework with stochastic job matching derived from Pissarides (1990), chapter 5, which is itself an extension of the basic search model of Diamond (1982). The novelty is that we explicitly consider an unemployment benefit b, and the budget constraint faced by the government which levies a pay-roll tax τ on workers and employers. Moreover, the benefit might be further constrained by incentive considerations.

The aim of this section is to establish a simultaneous equation system, which permits the determination of the endogenous unemployment rate u, the set of incentive compatible benefit levels, and the level of the pay-roll tax necessary to finance the latter. Stochastic job matching permits the derivation of a non-trivial wage distribution, which then may give rise to a non-trivial benefit distribution.

Assume that workers are identical ex ante, and if they search, they do so with the same intensity, but the productivity α of a particular job match varies. Its precise value is only revealed upon contact although the distribution of productivities $G(\alpha)$ is common knowledge. Workers are heterogeneous ex post. Since all workers are identical ex ante, they have the same reservation productivity α_r. If productivity has the distribution G with support $[\alpha_l, \alpha_u]$, then workers accept all jobs characterised by $\alpha \geq \alpha_r$:

$$\int_{\alpha_i}^{\alpha_u} dG = 1 - G(\alpha_r) \tag{3.1}$$

[2]Let u denote the unemployment rate and v the vacancy rate (being the number of job vacancies over the total labour force). Trade in the labour market is uncoordinated, time consuming and costly. This notion is captured by a matching function, $x(u, v)$, giving the fraction of job matches x as a function of the unemployment and vacancy rates. This function is commonly assumed to be homogeneous of degree one. Define $\theta := v/u$ as a measure of labour market

[2]Firms have a reservation productivity, but, as demonstrated below, the Nash bargaining rule governing wage determination implies that both workers and firms agree on a common reservation productivity.

tightness. Thus $x/v = x(\theta^{-1}, 1) =: x(\theta^{-1})$. The stochastic processes governing the economy are Poisson processes. A vacant job becomes occupied at a rate $q(\theta, \alpha_r) := [1 - G(\alpha_r)]\, x(\theta^{-1})$ since only jobs are formed which exceed the reservation productivity. Workers transit from unemployment to employment at rate $\theta q(\theta)$, but they become unemployed at the exogenously given separation rate s.

In equilibrium, the inflow into unemployment equals its outflow, $\theta q(\theta, \alpha_r)\, u = s(1-u)$, whence the Beveridge Curve (BC) in (v, u)-space is arrived at:

$$u = \frac{s}{s + \theta q(\theta, \alpha_r)} \tag{3.2}$$

The benefit system

A person may apply for a benefit b when unemployed, but the institutional conditions of eligibility may be more extensive. For instance, unemployment benefit may be paid for a limited duration only or it may be contingent on the contributions record of the claimant. Below two benefit schedules will be discussed, viz. a flat-rate (FR) and an earnings-related (ER) schedule. The attainment of any policy objective is, however, constrained by the scarcity of resources, and the social budget needs to be balanced. It is a common institutional practice that contributions are shared equally between employer and employee. Here it is implemented as a payroll tax τ on gross earnings w, levied in equal proportions on the two parties.

Firms

The firm has a standard neo-classical production function, exhibiting constant returns to scale, but because of the reservation productivity rule it has to be written as $\hat{F} = \hat{F}\left(K, N\alpha_f^e\right)$ with conditional expectations $\alpha_f^e = \mathcal{E}\left[\alpha | \alpha \geq \alpha_f\right]$, since firms have to forecast productivities. α_f is the reservation productivity of the firm below which workers are rejected. The production function can be re-written as $f(k)$, where $k := K/N\alpha_f^e$. The value of an occupied job J or a vacancy V are captured by asset value (or "no arbitrage") equations. The value of a job J is

$$r(J + \alpha k) = \alpha\left[f\left(k\right) - \delta k\right] - w\left(1 + \tau\right) + s\left(V - J\right) \tag{3.3}$$

since it produces $\alpha\left[f\left(k\right) - \delta k\right]$ but the firm has to pay a wage w and a tax $w\tau$, and loses a worker at the exogenously given separation rate s. r is the interest rate and δ the depreciation rate. The value of a vacancy V is

$$rV = -\Gamma + q(J^e - V) \tag{3.4}$$

where Γ is the search cost of the vacancy, and the vacancy becomes occupied at rate $q(\theta, \alpha_f)$. In equilibrium the value of the vacancy must be zero, $V = 0$, which implies $J^e = \Gamma/q$, since otherwise the firm would change its behaviour. The condition $J = 0$ yields the reservation productivity, since firms employ workers as long as the job is profitable, whilst reaping the surplus from the intra-marginal worker. In consequence,

$$\alpha_f = \frac{\left(1 + \tau\right) w}{f\left(k\right) - \left(\delta + r\right) k} \tag{3.5}$$

Firms choose k optimally, which implies $f'(k) = \delta + r$. After taking conditional expectations of the valuation of an occupied job J and imposing $V = 0$, the equilibrium condition on vacancy supply by firms becomes:

$$\alpha^e \left[f(k) - (\delta + r)\, k\right] - w^e (1 + \tau) - \frac{(r + s)\,\Gamma}{q} = 0 \qquad (3.6)$$

Workers

Being unemployed implies a value U to the unemployed because of the receipt of a benefit b and a chance of $\theta q(\theta)$ to become employed. The value of employment E is derived from a net wage $w\,(1 - \tau)$, but the worker may lose her job with an exogenous probability s. Consequently the asset value equations are :

$$\begin{aligned} rE &= w(1-\tau) - s(E - U) \qquad (3.7) \\ rU &= b + \theta q(\theta, \alpha_r)(E - U) \end{aligned}$$

Wages

The occupied job creates a surplus which must cover the search costs of both parties. It is commonly assumed in the search literature that the surplus bargained over by the worker and the firm is divided according to the Nash bargaining rule (Binmore, Rubinstein, and Wolinsky (1986)). This Nash rule implies that workers and firms have a common reservation productivity, $\alpha_r = \alpha_f$. Suppose that this surplus is divided equally [3]: the chosen wage then maximises

[3]Note that the resulting returns will, in general, not be efficient, since neither party obtains its respective marginal product.

$(E-U)^{0.5}(J-V)^{0.5}$, which implies the wage equation

$$2(1-\tau)w = b + [(1-\tau)\,/\,(1+\tau)]\,[\theta\Gamma + \alpha\,[f(k) - (\delta+r)k]] \quad (3.8)$$

A higher productivity is remunerated by a higher wage.

Flat-rate (FR) benefits and eligibility

If the benefit is a flat rate, it is most conveniently formalised as a constant fraction of expected wages $b = \lambda w^e$ where $\lambda \in (0;1)$. Following Pissarides (1990, p.99), it is also convenient to formalise the firms' search costs in a similar manner, $\Gamma = \gamma w^e$. If the social budget is balanced, expected expenditures have to equal expected incomes

$$\int_{\alpha_r}^{\alpha_u} (1-u)\,2\tau w d\tilde{G} = 2(1-u)\tau w^e = bu \quad (3.9)$$

where $\tilde{G}$ is the conditional productivity distribution. This formulation implies that all unemployed receive a benefit b, even those whose productivity falls below the generally accepted reservation productivity α_r. Solving (3.9) yields

$$\lambda = \left[\frac{1-u}{u}\right] 2\tau \quad (3.10)$$

Some algebraic manipulations yield the equation

$$(1+\tau) - \frac{1+\tau}{1-\tau}\frac{1-u}{u}2\tau - \gamma v\left[\frac{1}{u} - \frac{r+s}{s}\frac{1}{1-u}\right] = 0 \quad (3.11)$$

This is the so-called Vacancy Supply curve (VS), which, like the Beveridge curve, is usually analysed diagrammatically in (v,u)-space. Also an expression for the reservation productivity can be derived

$$\frac{\alpha_r}{\alpha^e} = \left(\frac{1+\tau}{1-\tau}\lambda + \theta\gamma\right)\left(2\left(1+\tau\right) - \frac{1+\tau}{1-\tau}\lambda - \theta\gamma\right)^{-1} \tag{3.12}$$

where α^e denotes the expected productivity.

In summary, the system consists of four equations, viz. (3.2), (3.6), (3.11), and (3.12). Given a pay-roll tax rate τ, the unknowns are u, v, k and α_r. Differentiating the VS curve (3.11)shows that, as usual, VS is upward sloping in (v, u)-space. As regards the Beveridge curve (3.2), the problem is more complicated because of the presence of α_r. But, following Pissarides(1990), making the assumption $(\partial\alpha^e/\partial\alpha_r)\,\alpha_r/\alpha^e < 1$ - at the optimum a rise in the reservation productivity increases the conditional mean proportionately less- the Beveridge curve can be shown to be downward sloping. This follows since the assumption implies that a change in the labour market tightness θ has a stronger direct effect on the probability of leaving unemployment, which exceeds the indirect effect through the reservation productivity. Finally, the unemployment rate is determined by the intersection of the two curves VS and BC depicted in Figure 1 , and k is derived recursively from $f'(k) = \delta + r$.[4]

The effect of a change in the pay-roll tax

An increase in the pay-roll tax may be examined diagrammatically, analysing the behaviour of (3.2) and (3.11) separately. Holding

[4]Observe that Pissarides'(1990) partial equilibrium model is nested within this general equilibrium framework, as can be seen by setting $\tau = 0$.

u constant and differentiating (3.11) totally gives an equation in $d\tau$ and dv

$$Ad\tau = \gamma \left[\frac{1}{u} - \frac{r+s}{s} \frac{1}{1-u} \right] dv \text{ where } A(\tau) := 1 - 2\frac{1-u}{u}\left(\frac{1+\tau}{1-\tau} + \frac{\tau}{(1-\tau)^2}\right) \tag{3.13}$$

At $\tau = 0$, $A(0) < 0$ and A falls monotonically with $\lim_{\tau \longrightarrow 1} A(\tau) = -\infty$. Thus $dv/d\tau < 0$ and VS shifts down. Concerning the Beveridge curve (3.2), the same reasoning which showed that BC is downward sloping leads to the conclusion that it shifts to the right. As τ increases both returns to the employed workers and to firms fall and the reservation productivity thus increases. In consequence, an increase in τ unambiguously leads to a higher unemployment rate. The size of the increase depends on the distribution of productivities G. This situation is depicted in Figure 1.

Flat-rate benefits and incentive compatibility

We have considered a (non-negative) value of the pay-roll tax below 100%, but the domain may be further constrained by considerations of incentive compatibility. Is the benefit sufficiently low that all unemployed beneficiaries have an incentive to search and to accept any given job offer ? The incentive constraint is

$$b = \lambda w^e \leq (1-\tau)\, w(\alpha_r) \tag{3.14}$$

since $w(\alpha_r)$ is the lowest wage in the economy, associated with the lowest admissible job productivity, viz. the reservation productivity α_r. Using the wage equation (3.8) and the equation for the reservation productivity yields

$$0 \leq \left[\frac{1-\tau}{1+\tau}\right]\theta\gamma w^{e} \tag{3.15}$$

which holds for all $\tau \in [0;1)$. The flat-rate benefit does not create an incentive problem because wages are always sufficiently high.

Earnings-related benefits (ER) and eligibility conditions

When the benefit is earnings-related, $b = \rho w$ with replacement ratio ρ, the inter-temporal structure of the economy becomes important since benefits are determined by past earnings. Assume then that agents are infinitely lived (an implicit assumption so far), but that benefits last for one period only so that persons continuously unemployed for more than one period are ineligible for the benefit. The balanced budget becomes

$$\int_{\alpha_r}^{\alpha_u}(1-u)\,2\tau w dG = \int_{\alpha_r}^{\alpha_u} u\rho w dG \tag{3.16}$$

since all those whose productivity falls below the reservation productivity, $G(\alpha_r)$, are not entitled to an unemployment benefit. In order to receive a benefit one must have been separated from the job at most in the last period. This is a realistic assumption because most unemployment benefit programmes (as distinct from unemployment assistance) make eligibility conditional on a work or contributions record [5]. Simplifying (3.16) yields an expression for

[5]Restricting benefit eligibility in this way is analytically convenient since "productivity in the last period" is the state variable in terms of which the subsequent

the replacement ratio ρ

$$\rho = \frac{1 - u_\rho}{u_\rho} 2\tau_\rho \tag{3.17}$$

(3.10) and (3.17) look similar but tax rates (and thus unemployment rates) may differ because of incentive problems. The subscripts have been added to emphasise the potential difference.

Earnings-related benefits and incentive problems

An incentive problem occurs in this regime if the person is entitled to a high benefit which exceeds a current low net wage offer. The benefit may be high because of a high previous productivity level α and a high pay-roll tax τ. Thus, to prevent this from happening, τ must be sufficiently low. How low ? Examining the wage equation (3.8), the lowest wage is achieved when $\alpha = \alpha_r$ and the person is

welfare analysis can easily be carried out diagramatically. Moreover, if $\tau \leq \tau^*$, the ER benefit is a mean-preserving spread of the FR benefit. However, the ER benefit differs from the FR benefit also in terms of its temporal nature. Suppose the eligibility restriction is removed so that the benefit depends on the wage of the last job. In this case the analytical details become more awkward. For instance, the balanced budget equation becomes contingent on the entire earnings history of the population. In order to satisfy the incentive compatibility constraint, τ has to be selected such that the highest possible benefit does not exceed the lowest possible wage offer. The former is attained when $\alpha = \alpha_u$ and the latter when $\alpha = \alpha_r$ in two consecutive periods. But although the eligibility restriction is removed, the logic of the subsequent welfare analysis remains unchanged (cf footnote 7). Therefore the restriction remains for expositional clarity.

not entitled to the benefit. The highest wage, and thus the highest benefit entitlement, is attained when productivity is at its highest, $\alpha = \alpha_u$, and the entitlement is in turn the highest [6]. Putting these together yields the incentive constraint

$$\rho \left[1 + \left(1 - \tau \frac{\alpha_u F + \theta \Gamma}{\theta \Gamma} \right) \right] \leq 2 \left(1 - \tau \right) \text{ where } F := f(k) - (r + \delta)\, k \tag{3.18}$$

Unfortunately one cannot derive a closed form solution since (3.18) depends on the unemployment rate u, which can be conveniently examined only diagrammatically as the intersection of the VS and the BC curves. However, examining the boundaries of (3.18) for $\tau = 0$ and $\tau = 1$ shows that at low levels of τ the constraint is satisfied, but at high levels the constraint is violated. By continuity there exists a critical level τ^*(when (3.18) holds with equality), below which tax rates are incentive compatible but above which they are not.

What happens if the eligibility rules are relaxed for the earnings-related benefit, so that receipt of the benefit is only contingent on being unemployed ? For instance, a previously ineligible unemployed person could receive a flat rate below the lowest earnings-related benefit. The following argument demonstrates that the critical level of τ, below which all tax rates are incentive compatible, exceeds τ^*

[6] These are given by $w_{\min} = \theta \Gamma / (1 + \tau)$ and

$$w_{\max} = \left[2 \left(1 - \tau \right) - \rho \right]^{-1} \left[\left(1 - \tau \right) / \left(1 + \tau \right) \right] \left[\alpha_u F + \theta \Gamma \right].$$

in this new regime. Set $\tau = \tau^*$. First observe that, if the unemployment rate is held constant, awarding previously ineligible unemployed persons a benefit increases the number of beneficiaries, which reduces the replacement rate. Now let u vary. The outside option for the previously ineligible person increases, raising labour costs. The reservation productivity and thus unemployment rise whilst the replacement rate falls further. As only more productive jobs are formed, the lowest offered wage has risen. The discrepancy between the highest benefit entitlement and the lowest offered net wage increases, the incentive constraint becomes a bit more relaxed, and the tax rate can be increased whilst remaining incentive compatible.

Summary

Figure 2 summarises the preceding discussion of the model in (u, v)-space. As long as $\tau \leq \tau^*$, the earnings-related benefit is incentive compatible, and the respective unemployment rates in the two benefit regimes are the same. But if the incentive compatibility constraint for the earnings-related benefit is violated, the unemployment rate increases and exceeds the one for the flat-rate regime. An increase in the pay-roll tax rate τ increases the unemployment rate u, but the precise increase depends on the distribution of productivities G.

What do the benefit schedules look like ? In Figure 3, they are depicted as a function of α, the last productivity of the unemployed claimant. It is assumed that $\tau \leq \tau^*$, so the unemployment rates associated with the two regimes are equal, which implies that the two benefit parameters satisfy $\lambda = \rho$. The diagram also shows the different eligibility conditions for the different regimes (but this argument will be generalised below). The position of the ER schedule depends on the location of α_r and two possibilities are arise: either α_r is sufficiently low, so that the ER schedule intersects the FR schedule, or it is so high that the ER benefit always exceeds the FR benefit.

3.3 Evaluating the performance of the benefit regimes

The criteria according to which the benefit regimes will be assessed are poverty, inequality, and a more general social welfare function. Do we obtain a universal ranking of benefit regimes or does the preference for one regime change when a different criterion is chosen?

These criteria incorporate value judgments and a certain degree of arbitrariness. For instance, not everyone may agree on the location of the poverty line. Or there may be disagreement about the sensitivity parameters of these criteria. Can this diversity of opinion be accommodated, so that the ranking of the benefit regimes does not change as these parameters change? Finally, to what extent does the ranking depend on the eligibility conditions?

3.3.1 Poverty

The properties of the conventional poverty indices are well known, but the choice of a particular poverty index may be quite arbitrary. The particular choice may be defended in the light of the special question posed, and for the present analysis the decomposable poverty index proposed in Foster, Greer, and Thorbecke (1984) is convenient. Let z denote an exogenous poverty line, F is the distribution of incomes y, and β a sensitivity parameter. The poverty index P_β only takes into account the income of the poor (all $y \leq z$), and weighs their (percentage) income shortfall from z , i.e. the gravity of poverty, by the sensitivity parameter β :

$$P_\beta = \int_0^z \left(\frac{z-y}{y}\right)^\beta dF(y) = \sum_{k=1}^{K} v_k P_{\beta k}, \text{ where } \beta \geq 0 \qquad (3.19)$$

The poverty index can be decomposed as follows. Partition the population into K groups with respective population share v_k, and let $P_{\beta k}$ denote the computed poverty index P_β for group k. Then the index is expressible as the weighted sum of poverty over the K subgroups of the population.

The natural partition in the model is to group the employed and the unemployed. If $\tau \leq \tau^*$, the population and the wages of the employed are the same for the two benefit regimes. In consequence, in an assessment of the relative performance of the two benefit regimes, one can concentrate on the poverty of the unemployed.

Does a change in the poverty line z change the ranking of the benefit regimes whilst keeping the sensitivity parameter fixed ? If

$\beta = 0$, then (3.19) becomes the head count index. If $z = z_1$ in Figure 3 then FR dominates ER but if $z = z_2$ then ER dominates FR. In fact $z = z_1$ is the trivial case since FR dominates ER for all sensitivity parameters β.

Does the ranking change when the sensitivity parameter β changes whilst the poverty line remains unchanged at $z = z_2$? Is there a trade-off between the incidence and the gravity of poverty ? For instance, two situations may emerge. In situation (a) a certain number of people live below the poverty line, but the income shortfall is not large. In situation (b) fewer are poor, but they suffer from a more severe income shortfall. Which situation is deemed worse is captured by the sensitivity parameter β. This trade-off can be examined when the poverty line exceeds the flat-rate benefit. A reversal of the initial ranking can be made by means of a continuity argument. If $\beta = 0$ then ER dominates FR, but if $\beta = \infty$ then FR dominates ER. Given the monotonicity of the poverty index, a critical level of β exists, β^*, such that for $\beta \geq \beta^*$ all FR dominate ER.

What happens to the ranking when eligibility conditions change[7]? Assume the earnings-related benefit is extended to cover all unemployed by awarding the previously ineligible a flat rate. In this case

[7] If the temporal eligibility condition is removed as suggested in the previous footnote, the logic of the above arguments remains unchanged. Mapping the incomes of the unemployed in (benefit,income)-space, the FR is a horizontal line which is cut by the ER benefit schedule. This crossing is sufficient to guarantee the existence of a critical β at which a reversal of the poverty ranking of the two regimes occurs (when the poverty line exceeds the flat-rate benefit).

the income shortfall of the poorest is less severe which translates into a higher critical level above which the ranking is reversed. On the other hand, FR may be restricted to the same beneficiary population as ER. Yet, the same continuity argument applies, leading to a reversal of the initial ranking.

3.3.2 Inequality

Another assessment criterion is income inequality. Similar decomposition considerations lead to the choice of the Generalised Entropy measure, defined by

$$GE_{\beta} = \frac{1}{\beta^2 - \beta}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\mu} - 1\right)\right]^{\beta} = \sum_{k=1}^{K} v_k^{1-\beta} s_k^{\beta} GE_{\beta k} + GE_{\beta B} \tag{3.20}$$

where β is a sensitivity parameter, y_i income of person $i \in \{1, .., n\}$, and μ average income. The index decomposes, so that $GE_{\beta k}$ measures inequality within group k, where v_k is its population share, and s_k its income share. $GE_{\beta B}$ measures the inequality between groups when each person within group k is assigned the average income of the group.

Which benefit system is associated with higher income inequality? The population partitions into the set of the employed, unemployed beneficiaries and unemployed non-beneficiaries. Let the eligibility rules for FR be (3.9) and for ER (3.16). If the pay-roll tax satisfies $\tau \leq \tau^*$, then the group size and the inequality of the employed are the same for the two benefit regimes. For FR, there is perfect equality amongst the unemployed beneficiaries. But with the eligibility rule (3.16) there is always a group of unemployed non-beneficiaries, whose income share is zero. This implies that FR always dominates ER, irrespective of β.

What happens if the eligibility rule (3.9) is changed so that the FR benefit covers exactly the same population as ER ? Choosing τ again in an incentive compatible manner, the various income groups have the same size. The between-group component, $GE_{\beta B}$, will also be the same, since ER then is a mean-preserving spread of FR. However, there is perfect equality amongst the group of beneficiaries when the benefit is FR. Thus, even under these new eligibility rules, FR always dominates ER.

3.3.3 Social welfare and work incentives: a trade-off ?

There might be an equity efficiency trade-off between social welfare and work incentives. Let the welfare criterion be the poverty index (3.19). A socially desirable pay-roll tax, then, is the solution to the programme $\min_\tau P_\beta$. The problem is not a trivial one, for although an increase in the benefit reduces the gravity of poverty, it might increase its incidence as the unemployment rate rises. Moreover, the initial benefit increase might be eroded away by a rise in the number of beneficiaries. Finally, is the socially desirable τ incentive compatible ?

If the benefit is a flat rate, differentiating λ with respect to τ yields $d\lambda/d\tau = [2/u]\,[(1-u) - (du/d\tau)\,\tau/u]$. The sign of this expression is ambiguous and depends on the elasticity of unemployment, and thus on the distribution of productivities G and the reservation productivity α_r . Two polar cases are imaginable: In case (a) the labour force is highly skilled (skill has to be loosely interpreted here since it is ex ante unobservable), where most frequency mass is concentrated on high productivity levels. A sufficiently small increase in α_r leads to a small increase in u and λ increases. In case (b) the labour force is badly skilled, and most frequency mass concentrates on low productivity levels. The same increase in α_r leads to a large increase in u and λ falls. If the poverty line z is sufficiently low, so that no employed workers are deemed to be in poverty, and after defining z relatively as $z = \pi w^e$, the social welfare criterion (3.19) reduces to $P_\beta = u\,[(\pi - \lambda)\,/\pi]^\beta$. As τ increases, so does u. Fewer persons are taxed at a higher rate, and the revenue is distributed amongst more persons. But in case (a) the increase in λ can outweigh the increase in u, an effect which becomes stronger the higher is β.

A further question is raised by the issue of incentive compatibility: is the socially desirable pay-roll tax incentive compatible ? For the flat-rate benefit regime this is trivially true, since all pay-roll taxes are incentive compatible. But for the earnings-related benefit regime, to which a similar analysis applies, there is a non-trivial incentive constraint. Whether the socially desirable τ satisfies this constraint

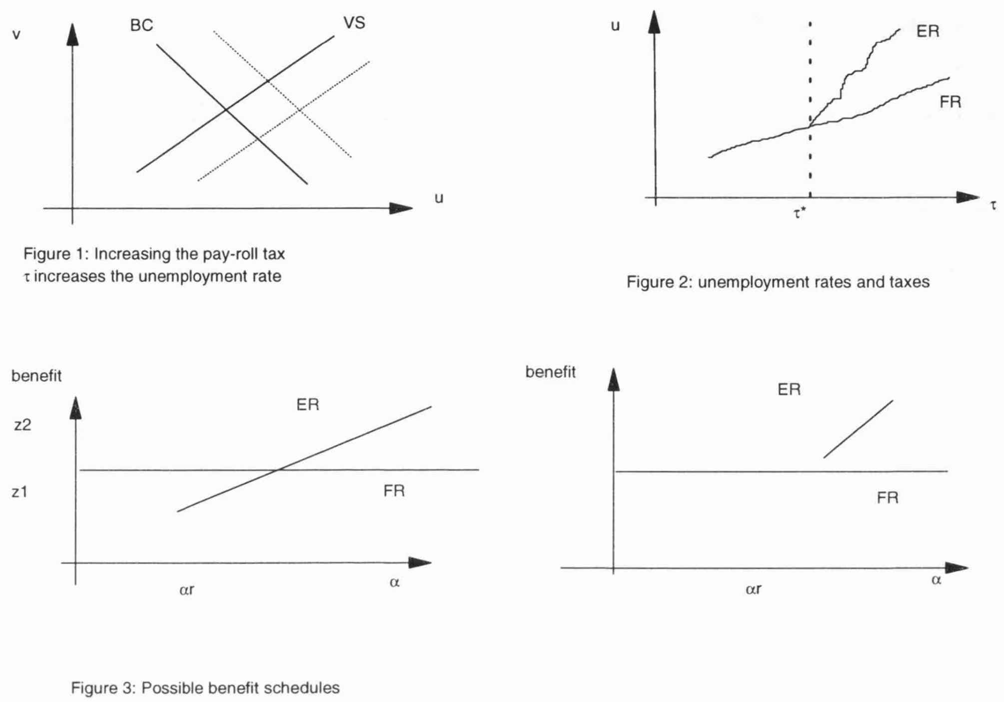


Figure 3.1:

depends again on the distribution of productivities G. The principal insight is, however, that the attainment of incentive compatibility is a value judgment which needs to be justified. An incentive compatible pay-roll tax might not be the socially desirable one. In particular it may be socially desirable that some low skilled persons face the wrong set of incentives, since the aggregate welfare effect exceeds the welfare loss caused by the latter.

3.4 Conclusion

Flat-rate (FR) benefits always produce lower inequality than earnings-related (ER) benefits[8], but poverty outcomes depend on the parameters of the poverty index. By changing these parameters, most initial rankings of the two benefit regimes can be reversed. Moreover, a trade-off between equity and efficiency might occur, which makes clear that the attainment of incentive compatibility for its own sake is a value judgment which needs to be justified.

However, one important assumption of the model is that agents are risk neutral, so that insurance has no role to play since agents only care about mean returns. This risk neutrality is a major cause for the negative results characterising the earnings-related benefit. Yet, insurance considerations are important in the design of actual tax-benefit systems. Finally, the normative question about the optimal insurance contract merits attention. If workers are assumed to be risk averse, this change destroys the linearity of the no arbitrage conditions for the worker and renders the model analytically intractable. All the same, some qualitative observations may be made, which point to ingredients a useful model of (social) insurance should incorporate. An earnings-related benefit seems to perform well with a proportional pay-roll tax since it achieves a desirable stabilisation of incomes in every state of the world. In the absence of incentive

[8] I.e. by removing deleterious incentive effects in benefits one can produce lower inequality.

considerations, the results of Yaari (1976) may be important who shows that the optimal consumption policy converges to mean consumption for an agent exposed to an iid income risk, when the rate of interest is zero and no borrowing constraints are imposed. However, Yaari does not justify this stochastic process by means of a labour market model.

3.5 Appendix: Risk aversion, tax-financed benefits, and Pareto improvements

This appendix attempts to indicate how the tax-benefit system can bring about Pareto improvements when agents are risk averse. The argument is only a partial equilibrium one, but the qualitative features are taken from the preceding (general equilibrium) model [9]. Assume that workers are homogeneous. The representative agent maximises the expected discounted stream of utility, where utility is solely defined over consumption $\mathcal{E}\left[\sum_{t=0}^{\infty}\beta^t U\left(c_t\right)\right]$, where $U(.)$ is increasing and concave, $U(0) = 0$ and $U'(0) < \infty$. With absent capital markets, the agent consumes all income in each period. As before, let s denote the exogenous job separation rate, τ the tax, b the benefit, and w the wage. w is drawn from the distribution F, with density f, over support $[0;\tilde{w}]$. For notational convenience, ignore τ and consider only b which may be zero. Below, we compare the welfare situation with no benefits ($b = 0$) to a situation with a small benefit ($b > 0$).

This problem will be analysed recursively. The agent's expected utility is, when accepting a wage offer w,

$U(w)+\beta\left[(1-s)\,v(w)+sv(b)\right]$, and if he chooses to search $U(b)+\beta\int_b^{\tilde{w}}v(x)f(x)dx$, where v denotes the value function. The latter

[9]This appendix is based on McCall (1970) and Lucas and Stokey (1989), section 10.7

becomes

$$v(w) = \max\left[U(w) + \beta\left[(1-s)\,v(w) + sv(b)\right]; U(b) + \beta\int_b^{\tilde{w}} v(x)f(x)dx\right] \tag{3.21}$$

which is well defined for the above problem. It is convenient to define $A := U(b) + \beta\int_b^{\tilde{w}} v(x)f(x)dx$, and it follows immediately that there is a unique w, w^*, such that $(1-\beta)A = U(w^*)$. This already is the optimal stopping rule: w^* is the reservation wage below which all job offers will be rejected. The value function then becomes

$$v(w) = \begin{cases} A \text{ for } w < w^* \\ \frac{U(w)+\beta sA}{1-\beta(1-s)} \text{ for } w \geq w^* \end{cases} \tag{3.22}$$

which is a continuous function at w^* and depicted in Figure 3.2 for the case $b = 0$.

In order to derive the defining equation of the reservation wage, eliminate A, and break up the integral from b to $\tilde{w}$ into one from b to w^* and one from w^* to $\tilde{w}$. It then follows, using $U(w^*) = (1-\beta)\,A$, that

$$\begin{aligned} & U(w^*)[1 + \beta s - \beta F(w^*) + (\frac{\beta^2 s}{1-\beta} + \beta)F(b)] \qquad (3.23) \\ = \; & (1 - \beta + \beta s)\,U(b) + \beta\int_{w^*}^{\tilde{w}} U(x)f(x)dx \end{aligned}$$

(Uniqueness of the reservation wage w^* due to a single crossing of the schedules can be verified by differentiating both sides with respect to w^*).

Analysing the effect of a 'small' benefit system reduces to examing $dv(w)/dw|_{b=0}$. What happens to the reservation wage defined by equation (3.23)? The value of the outside option rises whilst the value of the job falls if taxes are levied on the employed. The overall effect is an increase in the reservation wage. This is the incentive effect analysed in this paper. Since $U(w^*) = (1-\beta)\,A$, A rises as well. Whether the upper branch of equation (3.22) increases as well depends on the sign of the expression

$$-U'(w)\frac{d\tau}{db}|_{b=0} + \frac{\beta s}{1-\beta}U'(w^*)\frac{dw^*}{db}|_{b=0}.$$

If, for instance, no taxes are levied, then $dv(w)/dw|_{b=0} > 0$ unambiguously, otherwise the concavity of the utility function will play an important role.

For such a case, Figure 3.2 depicts the effects of introducing a 'small' unemployment benefit. Since $v(w)$ is the expected utility given the current state, the figure shows the areas for which the area is better off. It is not surprising that a small unemployment insurance benefit should make a risk averse agent better off; however, the figure shows that the incentive effect -an increase in the reservation wage- may, in fact, reduce the welfare of some agents.

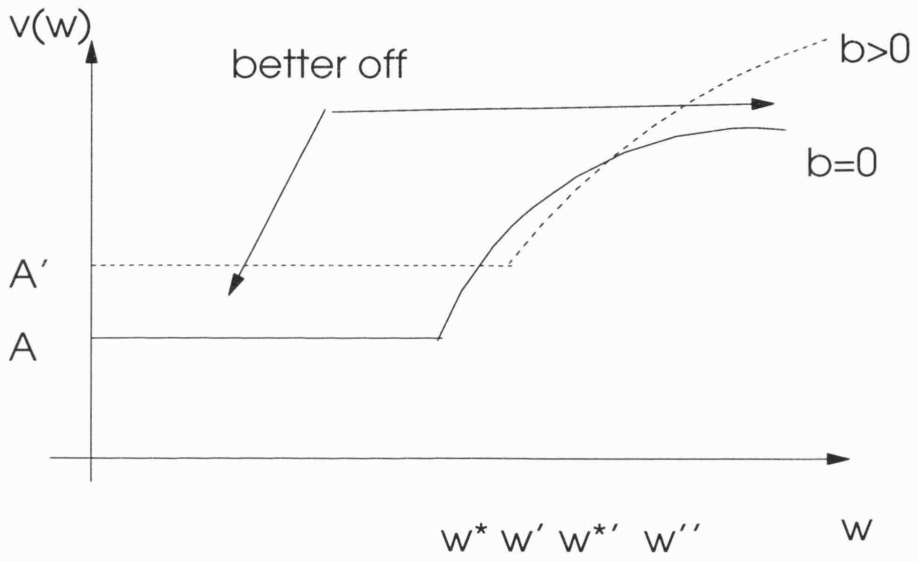


Figure 3.2: The value functions

Chapter 4

The statistical analysis of inequality and mobility indices

Abstract:We summarise results on the asymptotic distributions of standard inequality measures and derive a framework for making rigorous statistical inferences. A novel result is the derivation of the asymptotic distribution of the Shorrocks (1978) mobility index. Instead of using the (Gaussian) approximations, an alternative approach is to bootstrap the test statistics. The relative performance of these two approaches is assessed by comparing the lengths of the respective confidence intervals. Finally the robustness properties of the inequality indices are examined

4.1 Introduction

Inequality indices are statistical tools employed to measure 'inequality' within a distribution, but this measurement has normative implications, and these spheres may often conflict. As their normative properties are analysed extensively in Atkinson (1970), this papers focuses exclusively on their statistical properties. Moreover, a framework for making statistical inferences is established

An inequality or mobility index is a statistic T defined on a random variable income Y with unobservable distribution function F. The measure can often be written as a linear functional or the ratio of linear functionals over the distribution F, so the population measure can be denoted by $T(F)$. But since the researcher rarely has complete access to the population data, he is usually confined to a sample, and needs to estimate the true but unobservable distribution function F. The typical practice is to estimate the inequality measure non-parametrically by using the observable empirical distribution function $F^{(n)}$ as an estimate for F, and its derivative statistics (such as the sample mean or its order statistics). The value of the inequality measure is thus $T(F^{(n)})$.

However, the measure $T(F^{(n)})$ is itself a random variable. It has a sampling distribution which needs to be taken into account when interpreting a particular realisation t of the measure $T(F^{(n)})$. This issue is often ignored in the applied literature, when a particular value t is taken at face value. In such a case the measure is used as a descriptive device and not as a tool for rigorous statistical inference. In another situation, the researcher may have computed two different values of the inequality measure for different samples, this difference may not be statistical significant. Although the finite samples are not identical, they may come from the same distribution. Relying solely on the point estimate may induce him to draw a wrong conclusion.

$T(F^{(n)})$ is not the only conceivable estimator of the population parameter. Other non-parametric estimators of F can be used, such as the smooth bootstrap estimator $\hat{F}$, inducing the estimator $T(\hat{F})$. Several resampling plans will be examined.

The second problem analysed in this chapter is that of data contamination. The sample may be contaminated by data which is mistyped by the data provider, the typical example being a comma error. Or the surveyed person may have given a wrong answer. The researcher then does not calculate the inequality measure on the empirical distribution function of the true distribution F, but on a mixture distribution $(1-\varepsilon)\,F+\varepsilon H$ where H is the contamination and ε the proportion of the data affected. We address the question to what extent the inequality measures are robust to data contamination.

The chapter is structured as follows. In section 4.2, the asymptotic distributions of standard inequality and mobility measures are derived. A framework for making statistical inference is established, and procedures for dominance tests are discussed. Since these results are only asymptotic, section 4.3 examines the relative performance of the asymptotics to various bootstrap techniques when sample sizes are relatively small. by investigating which methods result in shorter confidence intervals. Section 4.4 examines the robustness of the standard inequality measures against outliers or departures from the hypothesised model.

4.2 Asymptotic methods when the empirical distribution function is used only once.

4.2.1 Asymptotic distributions of inequality and mobility measures

This section collects results on the asymptotic distributions of popular measures. The typical practice is that the researcher computes an inequality measure on the basis of the unmodified sample from the population. Let Y denote the random variable income with (unobservable) distribution function F and let $F^{(n)}(x) = (1/n) \sum I_{(Y_i < x)}$ denote the empirical distribution function putting equal weight on each observation where I denotes the indicator function and n is the sample size. The the measure is denoted by $T(F^{(n)})$.

The surprising result is that all the measures analysed here are asymptotically normally distributed. This result is astonishing since, in most cases, normality does not follow from a straightforward application of the central limit theorem. Indeed, the arguments involved to justify the claim are at times intricate and differ markedly from measure to measure. The result about the normality of the Shorrocks mobility index is new and supported by a complementary bootstrap analysis.

The Generalised Entropy index

Let GE_α denote the Generalised Entropy index with parameter $\alpha \neq 0, 1$. For these special cases de l'Hopitals rule needs to be applied before conducting the subsequent analysis. Then $GE_\alpha(F)$ is defined by

$$GE_\alpha(F) = T(\mu_1, \mu_\alpha) = \frac{1}{\alpha^2 - \alpha}\left[\frac{\mu_\alpha}{(\mu_1)^\alpha} - 1\right]$$

(in a slight abuse of notation) where $\mu_1 = \int ydF(y)$ and $\mu_\alpha = \int y^\alpha dF(y)$. α is a sensitivity parameter. The lower it is, the greater is the weight given to the bottom of the distribution. Using $F^{(n)}$ the sample estimator $\widehat{GE}_\alpha(F^{(n)})$ is readily computed to equal

$$\widehat{GE}_\alpha(F^{(n)}) = T(\hat{\mu}_1, \hat{\mu}_\alpha) = \frac{1}{\alpha^2 - \alpha}\left[\frac{\hat{\mu}_\alpha}{(\hat{\mu}_1)^\alpha} - 1\right]$$

where $\hat{\mu}_1 = \int ydF^{(n)}(y)$ and $\hat{\mu}_\alpha = \int y^\alpha dF^{(n)}(y)$. $F^{(n)}(y) \rightarrow F(y)$ uniformly in y by the Glivenko-Cantelli theorem, and $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_\alpha)$ is a consistent estimator of the population parameters $\mu = (\mu_1, \mu_\alpha)$. Of course, $n^{0.5}(\hat{\mu} - \mu)$ is asymptotically normally distributed with mean zero and covariance matrix Ω. $T(.)$ is a differentiable function and the delta-method can be applied (see Rao (1973, p.387)). It follows that

$n^{0.5}(T(\hat{\mu}) - T(\mu))$ is asymptotically normally distributed with zero mean and covariance matrix $\sigma^2 := T'_\mu \Omega T_\mu$, T_μ denoting the vector of derivatives. σ^2 can be consistently estimated as $\hat{\sigma}^2$ by substituting in the sample vector $\hat{\mu}$ for μ. Thus we obtain convergence to the standard normal distribution:

$$\frac{\sqrt{n}\left[\widehat{GE}_\alpha(F^{(n)}) - GE_\alpha(F)\right]}{\hat{\sigma}} \longrightarrow N(0;1)$$

The Gini coefficient

The theory needed to justify the normality of the Gini coefficient is much more intricate than in the preceding example because its usual estimator involves order statistics. But Hoeffding (1948) has developed a general theory of U-statistics, which he applied to the Gini coefficient in order to derive its asymptotic distribution. Let *Gini* denote the coefficient for the income population with distribution F and mean μ given by

$$Gini(F) = \frac{\delta}{2\mu}$$

where $\delta := \int\int |x - y| dF(x) dF(y)$. $Gini(F^{(n)})$ is bias-corrected so that

$$\widehat{Gini} = \frac{\hat{\delta}}{2\hat{\mu}}$$

where $\hat{\delta} := [n(n-1)]^{-1} \sum_{\alpha \neq \beta} |Y_\alpha - Y_\beta|$ is a consistent estimator of δ, and $\hat{\mu}$ is the sample mean. Since the income random variable is non-negative and assuming its second moment exists, Hoeffding's theorem applies and

$$\sqrt{n}[\widehat{Gini} - Gini(F)] \longrightarrow N(0;\Omega).$$

For the messy variance expression and its consistent estimator see Hoeffding (1948).

Lorenz curves

Beach and Davidson (1983), amongst others, derive the asymptotic normality of the Lorenz curve ordinates. Let ξ_{p_i} and p_i denote a quantile of the income variable and its population share, $p_i = F(\xi_{p_i})$, and if F is strictly monotone, $\xi_{p_i} = F^{-1}(p_i)$. A coordinate of the Lorenz curve is a pair $(p_i; \Phi(\xi_{p_i}))$ and the Lorenz curve consists of these ordered pairs where $i = 1, .., k$. The coordinate for the unobservable population distribution F is defined by

$$\Phi(\xi_{p_i}) = T(\xi_{p_i}; F) = \frac{1}{\mu}\int_0^{\xi_{p_i}} udF(u).$$

Substituting in the empirical distribution function $F^{(n)}$ yields the sample estimate

$$\hat{\Phi}(\hat{\xi}_{p_i}) = T(\hat{\xi}_{p_i}; F^{(n)}) = \frac{1}{\hat{\mu}}\frac{1}{r_i}\sum_{i=1}^{r_i} Y_{(i)}$$

where $r_{(i)} = [np_i]$ denotes the greatest integer less or equal to np_i and $Y_{(i)}$ is the i-th order statistic of the sample.

The key theorem to prove normality of the ordinate estimate is the fundamental theorem that the order statistics are normally distributed around their population analogues, and so are linear functions of the sample order statistics (see Rao (1973)). The principal requirement is that F is strictly increasing and twice differentiable. In consequence, letting $\hat{\Phi} = vec[\hat{\Phi}(\hat{\xi}_{p_i})]$

$$\sqrt{n}[\hat{\Phi} - \Phi] \longrightarrow N_k(0; \Omega)$$

For the messy covariance expression see Beach and Davidson (1983). Their principal merit is to propose a consistent estimator $\hat{\Omega}$ of Ω which does not require knowledge of F.

This result suggests an alternative non-parametric estimator of the Gini coefficient. Since the latter can be written as a linear function of Φ, say $1 - c'\Phi$ where c is a vector of coefficients, and as linear functions of normal variates are normally distributed, it follows that the Gini coefficient is asymptotically normally distributed with covariance matrix $c'\Omega c$.

The Shorrocks mobility index

The setting for a discussion of the Shorrocks mobility index is different from the preceding discussion, since it is defined on transition matrices of incomes.

Let $P = [p_{ij}]$, $i = 1, .., n$, $j = 1, ..., n$ denote the unobservable nxn stochastic transition matrix, satisfying $\sum_j p_{ij} = 1$, $i = 1, .., n$ and p_i the row vector $p_i = (p_{i1}, .., p_{in})'$. p_{ij} denotes the conditional probability of moving into state j next period, given that state i is occupied in the current period. The matrix of all cell counts is denoted by $X = [x_{ij}]$ and the row vector by $x_i = (x_{i1}, .., x_{in})'$, where x_{ij} is the number of observations falling into state j given state i.

Finally, let $n_i = \sum_i x_{ij}$ denote the total number of observations in each row i of X. The maximum likelihood estimator of the (first order Markov) transition probabilities is then defined by $\hat{P} = [\hat{p}_{ij}]$ with $\hat{p}_{ij} = x_{ij}/n_i$.

The large sample distribution of the Shorrocks index is based on well-known properties of the multinomial distribution (Kendall and Stuart (1977), p.381): it follows from the central limit theorem that x_i will tend, with increasing n_i, to have a $n-$variate normal distribution, with means $n_i p_{ij}$, variances $n_i p_{ij}(1-p_{ij})$ and $cov(x_{ij}, x_{ik}) = -n_i p_{ij} p_{ik}$ [1]. More specifically, $\sqrt{n_i}(\hat{p}_{ii} - p_{ii})$ will tend towards the normal distribution $N(0; p_{ii}(1-p_{ii}))$. Assuming the rows of P, i.e. the conditional distributions, to be independent, $trace(\hat{P}) = \sum_i \hat{p}_{ii}$ tends to $N(\sum_i p_{ii}; \sum_i p_{ii}(1-p_{ii})/n_i)$ and thus

$$\mu_1(\hat{P}) = \frac{n - trace(\hat{P})}{n-1} \rightarrow N\left(\frac{n - \sum_i p_{ii}}{n-1}; \frac{1}{(n-1)^2}\sum_i p_{ii}(1-p_{ii})/n_i\right) \tag{4.1}$$

Thus Shorrocks' index is asymptotically normally distributed. This result is verified by bootstrapping the mobility index 1000 times (using income data from the panel dataset described and extensively

[1] In fact, given this asymptotic normality, one can apply the delta method (Rao (1973)) to all standard mobility indices to demonstrate their asymptotic normality. The first order Taylor expansion of $\mu(\hat{P})$ about $\mu(P)$ is $\mu(\hat{P}) = \mu(P) + DM(P)(vec(\hat{P}' - P'))$ where $DM(P) := \partial DM(P)/\partial vec(P')'$. Since $\sqrt{n}vec(\hat{P}' - P') \rightarrow N(0, V)$ it follows that $\sqrt{n}(\mu(\hat{P}) - \mu(P)) \rightarrow N(0, \Sigma)$ where $\Sigma = DM(P)VDM(P)'$. See Trede (1995) for explicit derivations.

analysed in the next chapters). The result of this simulation is summarised by Figure 4.1 which depicts a histogram for the realisations of the Shorrock index generated in all 1000 bootstrap runs and shows the asserted normality.

4.2.2 A framework for hypothesis testing

Statistical inference: a difference of means test

Having derived the asymptotic distribution of the standard inequality and mobility measures to be normal, one can assess the statistical significance of a particular estimate by computing its standard error. Moreover, given two independent samples, one can address the question whether the computed measures are statistically significantly different. Since the asymptotic distribution of a standard measure is normal, the appropriate test is a difference-of-means test. Let $\hat{T}^{(i)}$ denote the sample measure, $\hat{\sigma}^2_{(i)}$ its estimated variance and $n_{(i)}$ the size of the sample drawn from populations $F_{(i)}$, $i = 1, 2$. The test statistic is

$$s = \frac{\hat{T}^{(1)} - \hat{T}^{(2)}}{\sqrt{\hat{\sigma}^2_{(1)}/n_{(1)} + \hat{\sigma}^2_{(2)}/n_{(2)}}}.$$

Under the null hypothesis that the two population measures are the same, s is asymptotically distributed as a standard normal distribution.

Given a chosen critical level, the significance level of the test, i.e. the probability of a type I error of wrongly rejecting the null hypoth-

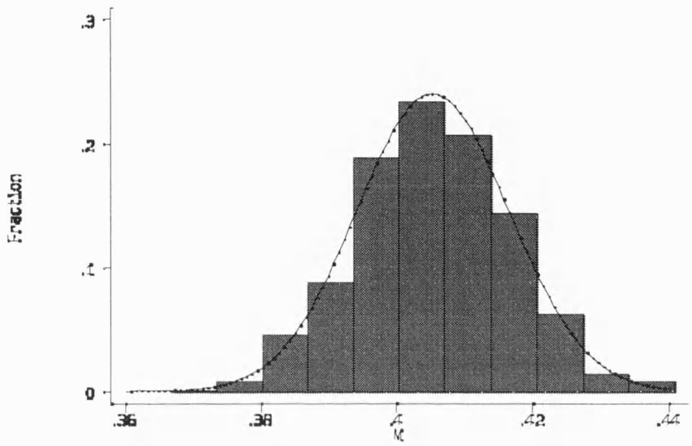


Figure 4.1: Histogram of the bootstrapped Shorrocks mobility index and the fitted normal distribution.

esis, is easily found by integrating the standard normal distribution. However, the power of the test, the probability of a type II error of wrongly accepting the null, depends on the difference of the two unobservable population measures $T^{(1)} - T^{(2)}$. If this difference is large, then the power of the test is large, but if it is small, so is the power of the test.

It is important to realise that this test requires independent samples. Whilst this may be safely assumed for geographically separated data, used, for instance, in international comparisons, this independence does certainly not hold for variables such as pre-tax pre-benefit income and post-tax post-benefit income for a given year. Two approaches may be pursued. First, if the type of correlation can be established, the inference rule might be adjusted. If the correlation is positive, the above test statistic s computed on the assumption of independence is lower than the true value s_{true}, $s_{true} \geq s$. The null hypothesis of same means is rejected if s is largel. Since s_{true} is even larger one can reject the null even for positively correlated samples. On the other hand, one cannot make an inference if s is sufficiently small as to suggest the acceptance of the null. In the case of negative correlation, the converse of above arguments applies.

The second approach is to attempt to estimate the covariance matrix for dependent samples. Davidson and Duclos (1995) achieve this aim for quantile-based test statistics, such as Lorenz curves. Let (Y, Z) be two jointly distributed random variables, F with inverse G denote the marginal distribution of Z, p a probability, and define the expectation $\gamma_p := E(Y|Z \leq G(p))$. Define a second set of random variables V and W analogously $(F^*, G^*, p', \delta_{p'} = E(V|W \leq G^*(p')))$. Then they derive a lengthy expression for the asymptotic covariance of the sample estimates $(p\hat{\gamma}, p'\hat{\delta})$ as:

$$
\begin{aligned}
& & E(YVI_{[0,G(p)]}(Z)I_{[0,G^*(p')]}(W)) - \\
E(Y|Z &=& G(p))E(VI_{[0,G(p)]}(Z)I_{[0,G^*(p')]}(W)) - \\
E(V|W &=& G^*(p'))E(YI_{[0,G(p)]}(Z)I_{[0,G^*(p')]}(W)) + \\
E(Y|Z &=& G(p))E(V|W = G^*(p'))E(I_{[0,G(p)]}(Z)I_{[0,G^*(p')]}(W)) - \\
pp'((\gamma_p - E(Y|Z &=& G(p)))(\delta_{p'} - E(V|W = G^*(p'))))
\end{aligned}
$$

Whilst the unconditional expectations can be readily estimated by their sample equivalents, Davidson and Duclos propose to estimate the conditional expectations by applying kernel density estimators. Moreover, distance estimators such as $\Gamma = [\Gamma_p] = [p(\gamma_p/\gamma_1 - \delta_p/\delta_1)]$are shown to be asymptotically normally distributed with covariance matrix of the form $J\Omega J'$ where J is the Jacobian of the transform Γ and Ω consists of the above covariances.

Dominance tests

The difference of means test can be employed to test whether two Lorenz curves derived from two independent samples are statistically significantly different. Since the sum of squares of independent standard normal random variables has a chi-squared distribution with degrees of freedom equal to the number of terms in the sum, it follows that the test statistic

$$c = \left(\hat{\Phi}_{(1)} - \hat{\Phi}_{(2)}\right)' \left(\hat{\Omega}_{(1)}/n_{(1)} + \hat{\Omega}_{(2)}/n_{(2)}\right)^{-1} \left(\hat{\Phi}_{(1)} - \hat{\Phi}_{(2)}\right)$$

has a chi-squared distribution with degrees of freedom equal to the number of population shares ($p_i < 1,\ i = 1,..,k$) at which the ordinates are estimated.

Moreover, these results can be used when designing a test as to whether Lorenz curves intersect, i.e. a test for Lorenz dominance. The test for Lorenz dominance is an example of multiple hypothesis testing (see Savin (1984)), since dominance is a joint statement about all individual Lorenz ordinates $\Phi(\xi_{p_i})$, $i = 1,..,k$. This set of individual hypotheses induces two types of "simple" tests. An intersection-union (IU) test is based on the union of all acceptance regions of each individual null hypothesis. If the multiple comparison null is $H_0 : H_0^{(i)} \quad [\Phi_{(1)}(\xi_{p_i}) \leq \Phi_{(2)}(\xi_{p_i})]$ for some i, and the alternative $H_A : H_A^{(i)} \quad [\Phi_{(1)}(\xi_{p_i}) > \Phi_{(2)}(\xi_{p_i})]$ for all i, then every individual null has to be rejected for the multiple comparison null to be rejected, and thus Lorenz dominance inferred. This induces the decision rule to infer that the first distribution Lorenz dominates the second if $\min(s_1,..,s_k)$ exceeds the critical level C_α (i.e. when H_A is accepted).

An alternative test is an union-intersection test (UI) which is based on the intersection of all acceptance regions of individual hypothesis. If the multiple comparison hypothesis is $H_0 : H_0^{(i)} \quad [\Phi_{(1)}(\xi_{p_i}) \leq$

$\Phi_{(2)}(\xi_{p_i})]$ for all i, and the alternative $H_A : H_A^{(i)} \quad [\Phi_{(1)}(\xi_{p_i}) > \Phi_{(2)}(\xi_{p_i})]$ for some i, then every individual null has to be accepted for the multiple comparison null to be accepted. The induced decision rule is to reject H_0 and infer that distribution one Lorenz dominates distribution two if $\max(s_1, .., s_k)$ exceeds the critical level C_δ . (This test could also be turned on its head by interchanging distribution one and two where the multiple comparison null is to be rejected if $\min(s_1, .., s_k) < -C_\delta$, so that Lorenz dominance is inferred when the first test is rejected but the second test is accepted.)

Which test is preferable ? The choice should clearly depend on both the significance level and the power of the test. The significance level of the IU rule is less than the significance level of the individual test, since the probability of wrongly rejecting the multiple comparison null when H_0 is true is bounded by the probability of wrongly rejecting an individual null when H_0 is true. In order to determine the significance level and thus the critical value C_δ of the UI test, the Bonferroni inequality is frequently used[2]. (Slightly sharper results can be obtained by the Sidak inequality or a further

[2]The Bonferroni inequality states that $\Pr\{A_1, .., A_p\} \geq 1 - \sum_{i=1}^{p} \Pr\{A_i^c\}$ where A_i is an event and A_i^c its complement. Applying this gives $\Pr\{\max\{s_1, .., s_k\} < C_\delta\} \geq 1 - k\delta$, where δ is the significance level of the individual test. If the multiple comparison test is to have significance level C_α, we require $1 - k\delta \geq 1 - \alpha$, which implies $\delta \leq \alpha/k$. The first inequality then determines the critical value C_δ. The Sidak inequality specialises to $\Pr\{\max\{s_1, .., s_k\} < C_\delta\} \geq (1-\delta)^k$ and a similar reasoning requires that $\delta \leq 1 - (1-\alpha)^{-k}$. Note that correlations may be arbitrary and that $(1-\delta)^k \geq 1 - k\delta$.

inequality derived from the studentised maximum modulus distribution.) But although the significance level may be bounded, the choice of the test should then be determined by its power. However, "little is known about the power of the Bonferroni test" (Savin (1984, p.860)). Howes (1993), comparing the two tests, also failed to impose a bound on the latter, but showed in a simulation study that the probability of inferring dominance when there is crossing can be high using the UI inference rule. He concludes that the IU should be used for inferring dominance, and the UI rule for inferring no dominance.

This method can also be employed to test for first order stochastic dominance, as carried out in Beach, Chow, Formby, and Slotsve (1994). Let $\gamma = [\gamma_i]$ denote the vector of cumulative means $\gamma_i = E[Y|Y \leq \xi_{p_i}]$ and $\mu = [\mu_i]$ the vector of quantile means $\mu_i = E[Y|\xi_{p_i-1} < Y \leq \xi_{p_i}]$. Beach and Davidson (1984) have shown $\hat{G} = (p_1\gamma_1; ...; p_k\gamma_k)$ to be asymptotically normally distributed with covariance matrix $\hat{\Omega}$. But since γ and μ are linearly related for equally spaced probabilities $\gamma = A\mu$, it follows that $\hat{\mu} = R\hat{G}$ is asymptotically normally distributed with covariance matrix $R\Omega R'$ where $R = (PA)^{-1}$ and $P = diag(p_i)$. The test for first order stochastic dominance then involves computing t-statistics for difference in sample means and then applying the union-intersection or intersection-union test.

4.3 Resampling Plans

Most inequality measures are linear functionals or the ratio of linear functionals of a distribution function F, $T(F) = \int t(y)dF(y)$. For instance, the function $t(.)$ may be as simple as $t(y) = y$ in the case of the mean. For the Gini, $\delta = T(F) = \int\int |x - y|dF(x)dF(y)$ or for the Lorenz ordinate $T(F) = \int yI_{(y \leq \xi_{p_i})}dF(y)$. Let F denote the true but unobservable distribution function. The linearity of the functional implies that one estimator of the population measure $T(F)$ can be computed by substituting in the empirical distribution function, $T(F^{(n)})$. This is the standard practice. However, several observations can be made at this point.

First, as regards standard practice, most researchers are content with a particular value $T(F^{(n)})$ and ignore the analysis of statistical significance. However, even if this analysis is conducted using the asymptotic theory of the previous section, sample sizes are often small after an extensive decomposition of the sample according to some characteristics. The question then arises whether the departure from the asymptotics is severe in small samples.

Second, $F^{(n)}$ is only one non-parametric estimator of F, and other estimates may be more suitable. In particular, if F is smooth and the sample size is small, it may be beneficial to use the smooth bootstrap estimator $\hat{F}$. Since $T(.)$ is linear, $T(\hat{F})$ is readily computed. These issues are analysed in the subsequent section.

The performance of the normal approximation may also be assessed by comparing its confidence intervals with those derived by applying bootstrap techniques. To this end a simulation study is carried out below.

4.3.1 The smooth bootstrap estimators

Let $\hat{f}_h$ denote the density estimate of the univariate distribution F with smoothing parameter h and kernel $K(.)$, given by

$$\hat{f}_h(y) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{y-Y_i}{h}).$$

Then

$$\hat{F}_h(y) = \int_{-\infty}^{y} \hat{f}_h(x)dx$$

is an estimator of the distribution function F. The functional $T(\hat{F}_h)$ is readily computed as

$$\begin{aligned} T(\hat{F}_h) &= \frac{1}{n}\sum_{i=1}^{n}\int t(y)\frac{1}{h}K(\frac{y-Y_i}{h})dy \\ &= \frac{1}{n}\sum_{i=1}^{n}\int t(Y_i+h\varepsilon)K(\varepsilon)d\varepsilon \end{aligned}$$

after the obvious change of variable with Jacobian h.

The resampling scheme here is to draw bootstrap samples from the estimate $\hat{F}_h$ instead from $F^{(n)}$. It is particularly attractive in small samples, since in this case ordinary bootstrap samples contain multiple occurrences of the same datapoint. However, this advantage vanishes the larger the sample is (see Hall, DiCiccio, and Romano (1989)). Sampling from $\hat{F}_h$ can be implemented in a straightforward manner. If the kernel $K(.)$ is Gaussian, one can set $X_i = Y_i + \varepsilon$ where ε has a standard normal distribution. Often it is advantageous to shrink $\hat{F}_h$, so that the modified data has the same variance as the original data (see Silverman and Young (1987)). Then set $x_i^* = \overline{y}^* + (1 + h^2/\hat{\sigma}^2)^{-0.5}(y_i^* - \overline{y}^* + h\varepsilon)$, where y_i^* are sampled with replacement from the original sample, $\overline{y}^*$ is its mean, $\hat{\sigma}^2$ its variance and ε is distributed as a standard normal. $(1 + h^2/\hat{\sigma}^2)^{-0.5}$ is the rescaling factor , and ε is Gaussian since the kernel is Gaussian.

A problem for all density estimation is the appropriate choice of the smoothing parameter as h offers a trade-off between smoothness and bias. In order to determine an appropriate bandwidth for the subsequent simulation study, two methods were applied to one random sample. First, the method of a reference distribution computes the optimal bandwidth if the sample were drawn from a normal distribution. This method yields h=.352. The second method is cross-validation, which is another attempt to minimise the mean squared error of the density estimate where the idea is to use the remaining observations to fit a density at each observation. This results in the estimate $h = .507$. For the subsequent simulation, a compromise value $h = 0.4$ has been chosen. Since the results of the first simulation using GE_2 were not encouraging, the bandwidth has been doubled, $h = 0.8$, for the subsequent simulations.

4.3.2 A simulation: Confidence intervals

In order to evaluate the relative performance of the normal approximation and the various bootstrap methods, the following simulation study is carried out. The first inequality index to be examined is GE_2, which is very quick to compute. Using $var(x) = E(x^2) - [E(x)]^2$, yields $GE_2 = 0.5var(y)/[\mu]^2$. Then the Gini and the Lorenz curve ordinate $\Phi(\xi_{0.5})$ of median income are tested. 1000 samples (500 for the smooth bootstrap simulations) of size n are drawn from $N(5;1)$ -the normal distribution with mean five and variance one. With F thus defined, the population indices are $GE_2(F) = 0.02$, $Gini$= 0.11284, and $\Phi(\xi_{0.5}) = 0.42$. In each iteration $B = 1000$ bootstrap samples are generated. Another simulation study is described below.

Let t denote the value of the estimator computed from the initial sample with empirical distribution function $F^{(n)}$. The bootstrap estimate $t^*_{(i)}, \quad i = 1, ..., B$ is computed by resampling with replacement from $F^{(n)}$. The nominal approximate coverage probability of the confidence interval is $100(1-2\alpha)$ and α is chosen as 0.025.

Confidence intervals for the normal approximation are based on the fact that $\sqrt{n}T(F^{(n)})$ is asymptotically normally distributed $N(T(F);\sigma^2)$. Standardising leads to the confidence interval $[t - 1.96\sqrt{\hat{\sigma}^2/n}; t + 1.96\sqrt{\hat{\sigma}^2/n}]$ where $\hat{\sigma}^2$ is a consistent estimator of the covariance of the inequality measure and n is the sample size.

Confidence intervals for bootstrap techniques are not unique.

Subsequently, three methods are implemented (see Shao and Tu (1995) or Davidson and Hinkley (1995) for a more thorough description). First, in the basic method the bootstrap distribution of $t^* - t$ is used to approximate the quantiles of the distribution $t - T(F)$. This yields the interval

$[2t - t^*_{((1-\alpha)(B+1))}; 2t - t^*_{(\alpha(B+1))}]$. Second, the percentile method assumes that there is an (unspecified) transformation of the estimator, yielding a symmetric distribution. Using the above quantile estimator and inverting the (unspecified) transformation gives the interval $[t^*_{(\alpha(B+1))}; t^*_{((1-\alpha)(B+1))}]$. However, since the assumed symmetry ignores bias and requires that $t^* - t$ and $t - T(F)$ have the same variance, the BC_α method, the third method, attempts to correct for bias and skewedness (see Efron (1987) for details).

4.3.3 Discussion

Table 4.1 reports the results; box plots are omitted for the sake of brevity. All methods produce confidence intervals which bracket the population value, and their lengths are comparable. The smooth bootstrap methods do not produce better results, i.e. shorter confidence intervals, than their competitors. The samples of size 100 appear to be sufficiently large to erode the advantage the method has in "small" samples. The (ordinary) bootstrap methods produce better results than the normal approximation, but the improvement is not dramatic. Surprisingly (but confirming results of other researchers

		n=100			n=1000		
	method	average lower	average upper	average length	average lower	average upper	average length
GE_2	normal approx.	.0144	.0257	.0113	.0182	.02184	.0036
	bootstrap						
	basic	.01426	.0255	.0112	.01819	.02183	.0036
	percentile	.01462	.0259	.01128	.01822	.02186	.0036
	BC_α	.01545	.0272	.01175	.0183	.02197	.0037
	smooth boot.						
	basic	.0135	.0262	.0127			
	percentile	.0141	.0267	.0126			
	BC_α	.0146	.0275	.0129			
Gini	normal approx.	.08489	.139	.05411	.1042	.12139	.0172
	bootstrap						
	basic	.09669	.1289	.03221	.1077	.1108	.0103
	percentile	.09499	.1272	.03221	.1075	.1178	.0103
	BC_α	.098	.1307	.0.327	.1078	.1182	.0104
	smooth boot.						
	basic	.0916	.1324	.0408			
	percentile	.0909	.1317	.0408			
	BC_α	.093	.133	.04			
LC	normal approx.	.408	.433	.025	.4164	.424	.0076
	bootstrap						
	basic	.407	.432	.025	.4163	.424	.0077
	percentile	.409	.434	.025	.4164	.424	.0076
	BC_α	.405	.431	.026	.4161	.4239	.0078
	smooth boot.						
	basic	.406	.436	.03			
	percentile	.4059	.4358	.0299			
	BC_α	.404	.434	.03			

Table 4.1: Simulation study: confidence intervals for sample sizes n. The population distribution is N(5;1). For each case 1000 simulations (500 in case of smooth bootstrap) were run, each comprising 1000 boostrap computations. Average lower (upper) refers to the lower (upper) bound of the confidence intervals, averaged across all simulation runs. Average length is the average length of the confidence interval.

in different applications, such as Burr (1994)) the computationally more complicated methods do not lead to shorter confidence intervals. In fact, the simplest method, the percentile method, performs best.

However, this simulation study has examined the simplest of situations in which samples are drawn from a normal distribution, i.e. a distribution which is smooth, unimodal and symmetric. In more complicated models the normal approximation may not perform as nicely. In order to test this possibility, two further simulations were conducted. First, samples are drawn from a Gamma distribution with shape parameter 2 in order to test the performance of the normal approximation on skewed distribution. Second, a bimodal normal mixture distribution is generated. Table 4.2 summarises the results. However, the relative performance of the normal approximation does not worsen, reaffirming the above conclusions.

4.4 Robustness

Very often, a hypothesised model fits the true but unobservable model only approximately; or the sample contains some contamination, such as outlier introduced by decimal point error in data transcription. The analysis of whether test statistics are robust against these types of model deviations is the subject of robust statistics, introduced into distributional analysis in Victoria-Feser (1993).

Formally, the observed distribution F_ε is a mixture between the

		G(2)			0.5N(2;1)	+	0.5N(5;1)
	method	average lower	average upper	average length	average lower	average upper	average length
GE_2	normal approx.	.174	.3268	.1528	.0963	.1720	.0757
	bootstrap						
	basic	.174	.324	.15	.0934	.1696	.076
	percentile	.177	.3275	.1505	.09866	.1748	.076
	BC_α	.1895	.351	.1615	.1024	.181	.0786
LC	normal approx.	.2066	.2723	.0657	.2448	.31354	.06874
	bootstrap						
	basic	.2023	.269	.0667	.2405	.31	.0695
	percentile	.2099	.2766	.0667	.248	.3179	.0699
	BC_α	.2016	.2687	.0.0671	.2376	.309	.0714
Gini	normal approx.	.2745	.469	.19	.2218	.378	.156
	bootstrap						
	basic	.3302	.4215	.09136	.2517	.35	.098
	percentile	.322	.4136	.09136	.249	.3477	.098
	BC_α	.3315	.4235	.092	.2575	.3585	.101

Table 4.2: Simulation study: confidence intervals. The first population distribution is Gamma with shape parameter 2. The second distribution is a mixture of normal distributions 0.5N(2;1) + 0.5N(5;1). For each case 500 simulations were run, each comprising 1000 boostrap computations . Average lower (upper) refers to the lower (upper) bound of the confidence intervals, averaged across all simulation runs. Average length is the average length of the confidence interval.

underlying distribution F and an error distribution H_z, $F_\varepsilon(x) = (1-\varepsilon)\,F + \varepsilon H_z(x)$, where ε is the proportion of the data affected. Using $H_z(x) = I_{(x \geq z)}$, a tool to measure locally the influence of a datapoint on the statistic $T(.)$ is the influence function (IF), defined as

$$IF(x;T;F) = \lim_{\varepsilon \longrightarrow 0} \frac{T(F_\varepsilon) - T(F)}{\varepsilon}$$

The crux is whether the influence of a small contamination or model departure can be so huge as to dominate the value of the statistic. In particular, if the IF is unbounded, $T(.)$ may be biased to the extent of being meaningless. Although H_z is a rather special distribution, the IF thus defined is sufficient to describe the maximal asymptotic bias of an estimator over a neighbourhood of the model, as the latter can be shown to be proportional to the IF.

The use of the IF extends, in fact, beyond the issue of robustness, and can be employed to derive the asymptotic distributions of section 2 for parametric cases. Supposing the distribution G is 'near' F, one can compute the von Mises/Taylor expansion of the functional $T(.)$. Then substituting in $G = F_n$ and invoking the Glivenko-Cantelli theorem,

$$T_n\left(F_n\right) = T(F) + \int IF(x;T;F)dF_n(x) + remainder$$

Evaluating the integral then yields

$$\sqrt{n}\{T_n - T(F)\} = (n)^{-0.5} \sum_{i=1}^{n} IF(X_i;T;F) + remainder$$

The first term of the right member is asymptotically normally distributed due to central limit theorem. One then needs to argue in the specific case why the remainder vanishes asymptotically. Finally the asymptotic variance equals

$$V(T;F) = \int IF(x;T;F)^2 dF(x)$$

In the following sections the influence functions of standard inequality measures are derived. It is shown that most of them are not robust against departures from the true model. Remedies to overcome this shortcoming are then discussed.

4.4.1 Robustness properties of quantile based inequality measures

In order to derive the influence functions of the class of quantile-based measures, it is necessary to derive the former for quantiles (this is done in Cowell and Victoria-Feser (1996b)).

Let F_ε denote the above mixture distribution with point mass contamination at z, $Q(F_\varepsilon;p)$ its p-quantile, and $Q(F;p) = x_p$ is the p-quantile of F. Using the definition of F_ε it follows that $Q(F_\varepsilon;p) = Q(F;[p - I_{(x_p \geq z)}\varepsilon]/(1-\varepsilon))$. If F is invertible at the quantile x_p, it is easily established that

$$IF(z;Q;F) = \frac{p - I_{(x_p \geq z)}}{f\left(Q(F;p)\right)}$$

For IF to be bounded requires, for $p \notin \{0,1\}$, $f(Q(F;p)) > 0$. At the boundaries, applications of de l'Hospitals' rule establish when IF is bounded.

The computation of quantile-based measures then involves the derivation of cumulative means, $\Phi(F;p) = \int^{Q(F;p)} xdF$. Using F_ε yields

$$\Phi(F_\varepsilon;p) = (1-\varepsilon)\int^{Q(F_\varepsilon;p)} xdF(x) + \varepsilon\int^{Q(F_\varepsilon;p)} xdH_z(x)$$

and differentiating this yields

$$\begin{aligned} IF(z;\Phi;F) &= -\Phi(F;p) + Q(F;p)f(Q(F;p))IF(z;Q;F) + z \\ &= pQ(F;p) - \Phi(F;p) + I_{(x_p \geq z)}(z - Q(F;p)) \end{aligned}$$

Thus a large z can render this IF unbounded, but if the income domain is bounded, so is, in general, this IF. In consequence, a measure like the generalised Lorenz curve is (in general) robust. However, most other quantile-based measures include a normalisation at the mean $\mu = E(x)$. But since the mean clearly is highly non-robust, these measures inherit this property.

4.4.2 Decomposable inequality measures

Another large class of inequality measures are the decomposable ones, which can be written in the form $I(f) = \psi[J(F,\mu(F)); \mu(F)]$ where $J(F;\mu) = \int \phi(x;\mu)dF(x)$. The generalised entropy indices constitute a subclass of indices. Since they depend on the mean and are linear in income, intuition suggests that they should fail to be robust. The subsequent analysis follows Cowell and Victoria-Feser (1996). The 'cleanest' result is obtained when the contamination is mean preserving and contains two points, resulting in

$$IF(z; I; F) = \frac{\partial}{\partial J}\psi(J; \mu)\left[-J(F; \mu) + 0.5\phi(x_1; \mu) + 0.5\phi(x_2; \mu)\right]$$

If the income domain is bounded, so is this IF. In this specialised computation, the mean remained unaffected. But since in general the mean changes as well, and it is clearly non-robust, the robustness properties of the index worsens too.

In a simulation study Cowell and Victoria-Feser (1996) compute the Theil index (GE_0) on a log-normal distribution $LN(1; 0.8)$. If 5% of the data is affected by decimal point errors, they find that the value of the index doubles.

4.4.3 Remedies

Although an inequality measure may fail to be robust but may have appealing normative properties, two approaches to 'robustify' the index can be pursued.

First, the sample might be trimmed, so that after removing the $[n\alpha]$ smallest and the $[n(1-\alpha)]$ largest observations, the statistic is computed on the remaining ones. This procedure leads to such concepts as a trimmed Lorenz curve

$$L_\alpha(F; p) = [\int_{Q(F;\alpha)}^{Q(F;1-\alpha)} xdF(x)/(1-2\alpha)]/[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xdF(x)/(1-2\alpha)]$$

Second, in the parametric approach, a parametric model for the income distribution is hypothesised, and its parameters are estimated in a robust fashion as suggested in the literature (Hampel, Ronchetti, Rousseeuw, and Stahel (1986)). Then, the inequality measures are computed using this parametric model. These computationally complicated procedures invariably trade-off efficiency and robustness.

4.5 Conclusions

This chapter has collected results which are important for the statistical analysis of inequality and mobility indices. The standard indices are shown to be normally distributed. Moreover, a framework for making statistical inferences has been established, and such tests as dominance tests were discussed. Although the derived results are valid only asymptotically, sample sizes above 100 appear to be sufficiently large, as bootstrap methods do not produces much shorter confidence intervals. Finally, the standard measures often fail to be robust to departures from the underlying model. Measures to robustify the index include restricting its domain, trimming, or parametric methods.

Chapter 5

The Panel Dataset GSOEP Described

Part III: The Empirical Analysis

Abstract: The panel dataset GSOEP is described, the tax-benefit simulations summarised, and data issues such as representativeness and sample selection are discussed.

5.1 Introduction

This chapter describes a panel dataset, the German Socio-Economic Panel (GSOEP), which is used in the subsequent empirical analyses of Part III of this dissertation. GSOEP exists in two incarnation, viz. GSOEP proper and the "Equivalent Datafile". The latter contains a very small subset of the variables contained in the former, but it includes some variables created by the data-provider, which are not contained in GSOEP proper such as an estimation of post-tax post-benefit household income. Its principal limitation is the length of the time series, encompassing currently the years 1984 to 1990 whereas GSOEP proper offers the latest survey responses; another drawback is the absence of some important variables. Chapter 8 uses and comments on the "Equivalent Datafile". In order to examine more recent data, GSOEP proper is used, but the tax-benefit system needs to be simulated. This chapter describes the raw data, summarises the necessary simulations and discusses data issues such as representativeness and sample selection. GSOEP proper is used in chapter 6 and 7.

5.2 The panel: description of the raw data

GSOEP is a clustered-sampled panel dataset which surveys panel members annually since 1984. Researchers outside Germany obtain a random draw of 95% of the gross sample housed at DIW (Deutsche Institut für Wirtschaftsforschung), a distribution method which is designed to preserve the anonymity of the sample members and to conform to the strict German data protection laws.

The data is collected by means of two types of questionnaires. The household questionnaire, to be filled out by the household head, collects information at the household level, such as housing costs and household asset income. The concept of the household refers to the dwelling, and therefore differs from the notion of the main tax unit or the core-family, because grandparents and other relatives may share the same dwelling. Each person aged at least 16 is surveyed in the individual questionnaire. Otherwise, parents answer the relevant questions about their children. Although the GSOEP is conceptually similar to the PSID, this interview procedure constitutes a major difference, since in the PSID only the household head is interviewed. The population is then partitioned into three samples. Sample A consists of West Germans only, sample B contains answer of foreign nationals. Since 1990 the panel has been extended to cover East Germany, collected in sample C. This conceptual organisation translates into wave92 containing six separate relevant data files. The provision of biographical information leads to a further supply of files, leading to a total of 17 files for wave92. The subsequent chapters concentrate on samples A and B only, because East Germany is still undergoing significant economic and institutional transformation.

The evolution of the panel is determined by the panel rules governing the treatment of a person. The principal concept is that of an original GSOEP member (as distinct from a household), being a person who has been surveyed in 1984 or who has been born to one. They remain members of the panel, even if they leave the original household. On the other hand, a person who joins the household of an original GSOEP member is interviewed as long as the person remains associated with the latter. Since 1988 this follow-up concept has changed in that all new persons, having gained an association with an original GSOEP member, are kept in the panel, even if this association later dissolves. At its inception in 1984, 16205 persons were registered. In 1992 65% of original panel members remained. This attrition is counterbalanced by the entry of new persons, who become associated with an original member. The consequence is twofold. The total number of interviews carried out has stabilised, but the dataset has ceased to be rectangular.

GSOEP, of course, contains a wealth of income variables which, unfortunately, cannot be used as immediate input into the empirical investigations. The subsequent chapters, for instance, seek to analyse the welfare properties of the income distribution or the distributional consequences of the German tax-benefit system. To these ends, the concept of post-tax post-benefit income is the most appropriate, but in this respect the dataset is most user-friendly. On the one hand, the household questionnaire includes a question about the household head's estimate of the household's disposable income. Unfortunately such an approach is notoriously flawed, and Rendtel, Langeheime, and Bernsten (1992) show that 50% of the sample dramatically underestimate this quantity. On the other hand, pre-tax variables proliferate and the only solution is to accept the fate lamented by Johnson in chapter one, and to pursue the unthankful task of simulating the tax-benefit system.

5.3 The derived data, simulations, and income definitions

Since most data about incomes are provided as gross amounts, an extensive tax-benefit simulation has to be applied in order to arrive at (post-tax, post-benefit) disposable income. The salient features of the (non-integrated) German tax-benefit system are described in Schluter (1995), which summarises the algorithms implemented in this simulation and we just summarise the most prominent institutional features. All benefits are tax-free, except for pensions. Benefits derived from social insurance claims are mostly earnings-related, whilst most assistance programmes requiring a means-test pay flat rates. The financial burden of child-rearing is redistributed both by means of direct and indirect benefits, viz. a universal child benefit, special means-tested programmes, and universal tax allowances. Married couples use the method of split taxation, in which joint taxable incomes are notionally divided equally before determining the applicable marginal tax rate. This method translates into a smaller marginal tax relative to individual assessment, if an earnings difference between the partners exists. The marginal rates of income tax are linearised, ranging from 19% to 53%.

The concept of the surveyed household refers to the dwelling, and therefore differs from the notion of the main tax unit or the core family, because grandparents and other relatives may share the same dwelling. An automatic procedure has been applied to identify the members of the core-family, i.e. the principal tax unit. Whenever possible, the method of split taxation is applied to a married couple. Otherwise the person's tax liability is assessed individually.

The simulation of social security contributions has been based on some simplifying assumptions. Privately insured persons have been treated as if they were insured by the relevant statutory insurances. This approximation is quite reasonable, given that, for instance, 90% of the population are insured by the statutory health insurance. Moreover, only the very rich can opt out of some statutory programmes, and then this choice reflects preferences and consumption aspects. Second, German civil servants (Beamte) enjoy special privileges and a number of fringe benefits. In the simulation all 310 cases (wave92) have been treated like ordinary workers, paying, as a first approximation, the same social security contributions as everyone else.

Finally, since the data for wave 92 provides only an indicator for the receipt of social assistance, it has been stipulated that the entire household received this means-tested benefit for the whole

items	income label
gross earnings (including bonuses and perks) + gross pensions + maintenance payments	annual gross income
annual gross income - social security contributions - taxes + benefits	annual disposable (post-tax post-benefit) income

Table 5.1: Income Definitions

year, amounting to the standard rates. Actual one-off payments could therefore not be included.

Table 5.1 summarises the income definitions used in the subsequent analysis.

The following income concepts are employed: annual pre-tax pre-benefit income includes gross earnings, gross pensions and maintenance payments received. In particular, bonuses and perks are taken into account. All other social security payments are classified as benefits. Deducting social security contributions, tax and tax-surcharges yields post tax income. Adding the non-taxable benefits post-tax, post-benefit annual income is arrived at.

The household questionnaire also provides a question about financial wealth and investment income, viz. about interest, dividends, and rent received. However, these items have not been included in the income definition employed here for several reasons. Given that the household is not necessarily identical with the core family, it is not clear why the household head should possess all the relevant information about the financial wealth of all the household members.

Second, the item suffers from a low response rate. The problems associated with this variable may be gathered from the fact that a separate one-off investment survey suffered not only from a low response rate, but also led to a permanent 5% attrition for the entire panel [1].

In order to take into account the economies of scale within the families, its size and composition, the McClements' equivalence scales are applied. Although the choice of any equivalence scale is arbitrary, the McClements' scale is widely used amongst researchers and the UK Department of Social Security [2].This choice has been made for several reason. First, Burkhauser et al.(1994) argue that the German social assistance scale implies too low scale economies. Second, using the McClements' scale permits a first comparison with comparative results for the UK, such as found in Jenkins (1994) [3]

The net sample of wave 92 used in this next chapter has been arrived at after dropping selected observations. East Germans were

[1]Ignoring asset income, however, also introduces a bias, as, for instance, a sizesable fraction of the elderly rely more heavily on this income source. The unobserved income distribution for the elderly may thus have longer tails, resulting in greater inequality amongst them than the analysis in chapter 6 suggests.

[2]Cf. DSS(1993) "Households below average income", for instance. Banks et al.(1993) provide a recent appraisal of the McClements scale.

[3]The results of Buhmann et al.(1988) and Coulter et al.(1992) demonstrate that the evaluation of income inequality can significantly change when different equivalence scales are applied. In fact the latter discover a U-type relation between the value of the inequality index and the elasticity of the equivalence scale.

identified and filtered out, because East Germany is still undergoing significant economic and institutional transformation. Then all households living in a community home have been dropped following the precedent established by the PSID, because the inspected data looked too peculiar. Finally, after performing all the computations, all persons were dropped whose disposable income fell below the (arbitrary) cut-off point of DM1000. To get an idea of the various sample sizes, the resulting net sample for wave92 contains N_{92}=10107 observations in H_{92}=3822. All other waves received a similar treatment, resulting in sample sizes above 9000. For the analyses of income dynamics incomes are evaluated at 1992 prices.

5.4 Representativeness

This issue of representativeness [4] is clearly important, since the subsequent chapters seek to draw inferences about the (West German) population and not be confined to making statements about one unique sample. Table 5.2 juxtaposes some characteristics of the net sample with census data pertaining to West Germany. Two observations can be made. The sample is 'close' to the population but the data needs to be weighted wherever possible. This is not unexpected since various groups have different sampling probabilities and

[4] On the issue of panel attrition see Rendtel(1995). The problem of weighting is extensively discussed in Pischner et al.(1995), and Riebschläger(1995) juxtaposes the weighting procedures of GSOEP, PSID and the BHPS.

wave 92	population	unweighted sample	weighted sample
unemployment rate [%]	6.3	5.87	4.9
active workforce [%]	46.5	46.6	45.4
foreign households [%]	9.5	28.4	6.3
source: own calculations; Datenreport (1994) and Jahrbuch (1994)			

Table 5.2: Representativeness of the sample

foreigners are deliberately over-sampled.

Moreover, the sampling probability of any wave is determined by three components: first, the probability of being sampled in the first wave. Second, the follow-up procedure needs to be taken into account, since new panel members already had a chance of being sampled in the first wave. Moreover, the different drop-out probabilities of different socio-economic groups can be controlled for. Table 5.1 illustrates this problem of panel attrition and quantifies its causes. However, the recent large influx of immigrants is not controlled for, resulting in a severe under-representation of this group.

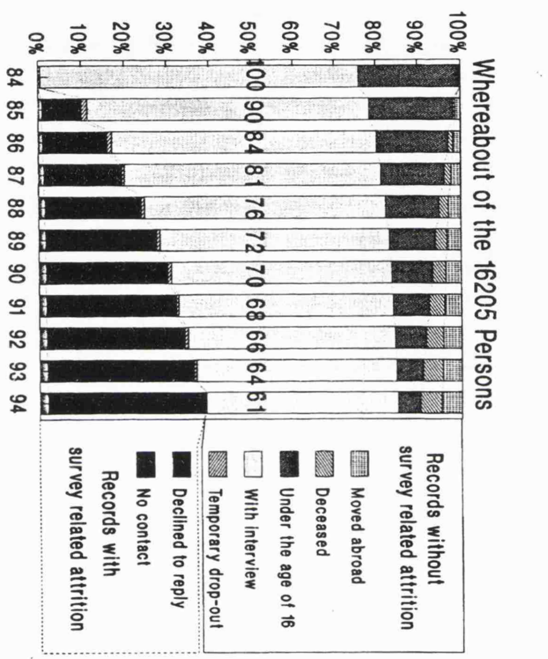


Figure 5.1: Panel attrition for GSOEP 1984-1994. Source: Rendtel (1995)

Chapter 6

Income Distribution and Inequality in Germany

Abstract: Kernel density estimates are employed for an exploratory analysis of the distributional consequences of the German tax-benefit system using GSOEP. The equalising forces of the system are examined for the year 1991. Moreover, the anatomy of the remaining income inequality -i.e. of post-tax post-benefit income- is thoroughly examined.

6.1 Introduction

The two preceding chapters examined some of the essential prerequisites for a rigorous empirical analysis. In this chapter we return to the main theme of this dissertation -the importance of intermediate institutions for the distributional outcome. Taking its cue from chapter 3, we focus again on the tax-benefit system which transforms the original distribution of factor incomes and results in the final distribution of (disposable) income amongst persons. Or, as Atkinson (1996) observes: "The gross incomes generated by production are typically modified by taxation, used to finance public spending, including transfers which constitute [another] source of personal incomes." Using the German panel dataset GSOEP we examine in detail the distributional consequences of the German tax-benefit system for the 1991 cross-section and thereby describe the empirical transformation of the distribution.

Although the tax-benefit system may, in the aggregate, constitute an equalising force, different subgroups of the population may fare differently. The final distribution may be a mixture distribution. The next natural step then is to investigate the anatomy of the resulting (disposable) income distribution. We examine to what extent the distributional implications of the German tax-benefit system vary systematically with personal characteristics, such as occupational status, work status, age and nationality.

These questions are addressed using two complementary approaches.

First, since one of the guiding questions of distributional analysis is "What does the income distribution look like ?", the non-parametric technique of kernel density estimation is employed in order to estimate and depict the relevant distributions directly. Whilst not being without its own problems, this method avoids some heroic distributional assumptions of parametric methods, and appears natural for an exploratory analysis of income distribution. To the end of interpreting their welfare properties conventional tools such as inequality indices and Lorenz curves are applied.

The structure of this chapter is as follows. Data issues and income definitions are discussed in the preceding chapter. Section 6.2 briefly summarises the method of kernel density estimation and outlines a bootstrap test for multimodality. The distributional consequences of the German tax-benefit system are analysed in section 6.3 by directly estimating the income distributions. Their welfare properties are assessed by testing for Lorenz dominance, stochastic dominance, and the computation of standard inequality indices. Each approach highlights different aspects of the equalising forces of the German tax-benefit system. In Section 6.4 we analyse the underlying structure of the remaining inequality by partitioning the population according to certain characteristics. The distribution of disposable income is shown to be a mixture of different underlying distributions. The diagrammatic interpretation is confirmed by decomposing the generalised entropy index. Section 6.5 concludes.

6.2 Estimating densities and assessing their properties

This section summarises the principal statistical methods used in this chapter. Section 6.2.1 is concerned with kernel density estimation and section 6.2.2 discusses a test for multimodality. Since chapter 4 collects results on the sampling distribution of inequality indices and establishes a framework for statistical inferences, these results will not be repeated here.

6.2.1 Kernel density estimation

Let x be the income random variable with distribution F and corresponding density f. X_i denotes a datapoint of the sample of size n. The kernel density estimator [1] $\hat{f}_n(x)$ estimates the density f at the point x and has the general form

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i}^{n} K\left(\frac{x - X_i}{h}\right)$$

where h represents the bandwidth around the point x, which may depend upon the number of observations n. The estimator can be visualised as imposing a window of length h about the point x and

[1] Silverman (1986) gives a concise introduction to density estimation and a broad overview of the pertinent issues. Prakasa Rao (1983) is more technical. A more recent review is carried out in Härdle and Linton (1994). See Deaton(1989) for a discussion extolling the merits of these non-parametric methods in analyses of large household survey datasets.

aggregating all datapoints X_i according to rule $K(.)$. The kernel used in this chapter is chosen with reference to the global accuracy of the induced estimator $\hat{f}_n(x)$, measured by the mean integrated square error. The smoothing parameter h can be chosen to minimise this objective function, but, inevitably, the optimal h also depends on the unknown f. A well known kernel is the ('optimal') Epanechnikov kernel given by:

$$K_e(t) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}t^2) & if \quad |t| \leq \sqrt{5} \\ 0 & otherwise \end{cases}$$

The smoothing parameter is the second choice to be made. The problem of choosing the smoothing parameter is illustrated in Figure 6.1 and Figure 6.3 , in which the density of one variable has been estimated using two different widths. The larger width irons out some of the bumps created by the smaller width, and by over-smoothing it may obscure some of the finer details of the underlying true density.

There is a large theoretical literature on choosing h under various conditions ranging from the quite arbitrary and informal method of fitting by inspection to highly sophisticated automatic algorithms. The applied literature reflects an absence of inexpensive automatic selection programmes. Deaton (1989), for instance, uses the informal method of inspection, whereas Quah (1994) uses the method of a reference distribution: the smoothing parameter is derived as the optimal width minimising $MISE$ if both data and kernel were Gaussian. It has been noted (Silverman(1986), p.46), however, that this method over-smooths multimodal or highly skewed densities because the width is too wide. Cowell et al. (1994) use an adaptive width, which follows similar ideas resulting in $h = 0.9An^{1/5}$ where A:= min(sample standard deviation, interquartile range/1.34) for the pilot bandwidth as given in Silverman (1986, p.48).

In this chapter the width is fitted by inspection and the following algorithm is applied: the starting point is a small width producing an erratic density estimate. Then, the width is gradually increased until a smooth estimate is arrived at (see Figure 6.1). Cross-checking, using the same technique, is done by applying the above algorithm both to the log-transformed data and to the re-transform (see Figure 6.3). The latter step yields a density estimate with an exponentially increasing bandwidth which accommodates the sparseness of the data in the right tail of the (skewed) income distribution.

6.2.2 A bootstrap test for multimodality

The test, proposed in Silverman (1981), is based on the fact that the number of modes of a Gaussian density estimate is non-decreasing as the bandwidth h increases. For a test of k-modality, the null hypothesis is that the number of modes equals k, against the alternative hypothesis of more than k mode. The test invokes the theory of bootstrapping. It is convenient to rescale the density, so that the modified estimate (the 'smooth bootstrap') has the same variance as the original sample. The key idea is to draw bootstrap samples x_i^* from the smooth bootstrap and to examine the number of modes

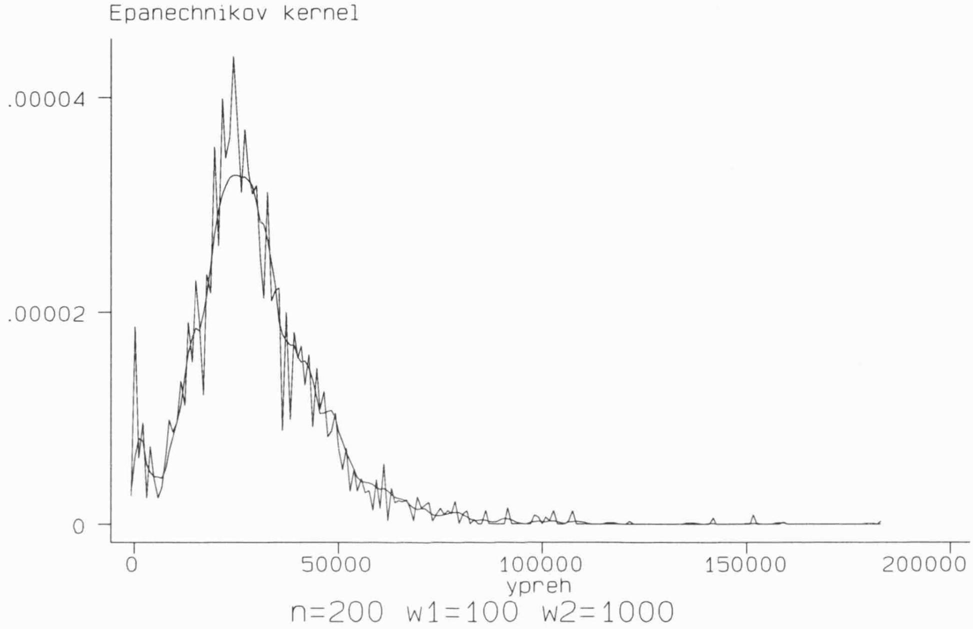


Figure 6.1: The importance of the smoothing parameter h: two density estimates at n equidistant points with width w.

of its distribution. This sampling can be directly implemented by setting

$$x_i^* = \overline{y}^* + (1 + h^2/\hat{\sigma}^2)^{-0.5}(y_i^* - \overline{y}^* + h\varepsilon)$$

where y_i^* are sampled with replacement from the original sample, $\overline{y}^*$ is its mean, $\hat{\sigma}^2$ its variance and ε is distributed as a standard normal. $(1 + h^2/\hat{\sigma}^2)^{-0.5}$ is the rescaling factor , and ε is Gaussian since the kernel is Gaussian. Let h_k denote the smallest value of h producing a k-modal density, and h_k^* its equivalent for the bootstrap sample. The p-level is approximated as a quantile of the h_k^* distribution by $\#\{h_k^* > h_k\}/B$, B being the number of bootstrap samples.

The test has the following interpretation. A large value of h_k, indicating that a lot of smoothing is needed to generate k modes, is taken as evidence against the null hypothesis of $k-$modality. As a consequence, a large value of p is taken to support the null hypothesis, whilst a low p-value constitutes considerable evidence against it[2].

[2]In the computer implementation a mode is defined as a point at which the slope of the density changes sign. In order to avoid contamination from noise in the tails of the distribution only point estimates above a critical level were considered. The robustness of the results with respect to these critical levels was also examined, and the computer implementation was tested on various model densities.

6.3 The estimates: The distributional consequences of the German tax-benefit system

The distributional consequences of the German tax-benefit system are best investigated directly by examining the kernel density estimates of the relevant income distributions, depicted in Figure 6.2. The net distribution has a large mode around the median, whilst the gross distribution is more spread out over its support. In fact, the Silverman (1981) test supports the hypothesis of a unimodal net distribution. The approximate p-value is calculated to be 0.78 and the critical bandwidth was found to be 1273.07; the former declines to 0.418 and the bandwidth to 1207.52 in a test for bimodality. Hence there is statistically significant evidence that the distribution is unimodal (but observe that both critical bandwidths lie close together). In contrast to this unimodality, net income distributions in the US and the UK have undergone a process of polarisation, resulting in bimodal shapes (see Burkhauser, Crews, Daly, and Jenkins (1995), and Jenkins (1995))

The distribution of disposable income is again depicted in Figure 6.3. Its upper panel provides an estimate of the log-transformed data with two different smoothing parameters. This transformation permits a nice representation of the tail of the distribution without obscuring detail at the lower tail. Retransforming the abscissa values again in the lower panel of this figure yields an increasing width. However, the principal features of the initial density estimate remain intact.

The effects of the tax-benefit system are intuitive: the lower tail of the gross income distribution is squeezed to the right as benefits become available to persons on low incomes, whilst higher marginal taxes pushes the upper tail of the income distribution to the left. These equalising forces confirm the common wisdom of Germany being a relatively equal society.

The welfare properties of these income distributions can be assessed by computing several inequality indices and other summary statistics, reported in Table 6.1. Mean and median of the net distribution do not lie too far apart, the median being about 91% of the mean. This already suggests that the post-tax, post-benefit distribution is relatively equal. The inequalities of both gross and net income distributions are measured by means of the Gini coefficient and the generalised entropy measure GE_{α} with sensitivity parameter $\alpha \in \{-1,0,1\}$.

All point estimates register the equalising forces of the tax-benefit system, which a standard difference-of-means test (see chapter 4) confirms as being statistically significantly.

Given these unambiguous rankings, the next natural step is to test for Lorenz dominance. Table 6.2 and Figure 6.4 confirm that the net distribution Lorenz indeed dominates the gross distribution -the above inequality indices inevitably produced an unambiguous rank-

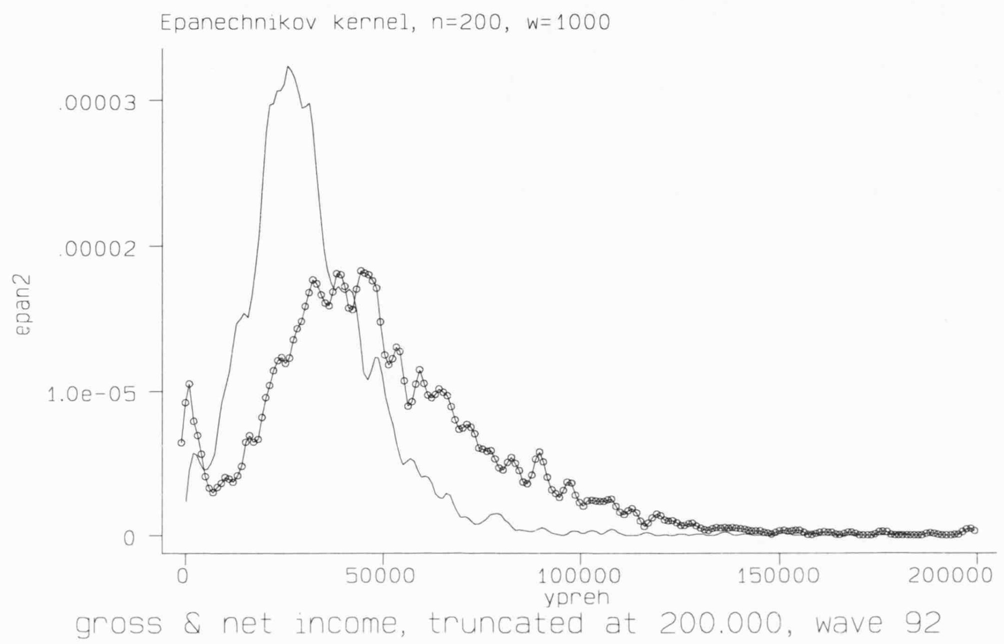


Figure 6.2: Density estimates. Note: circles refer to gross income

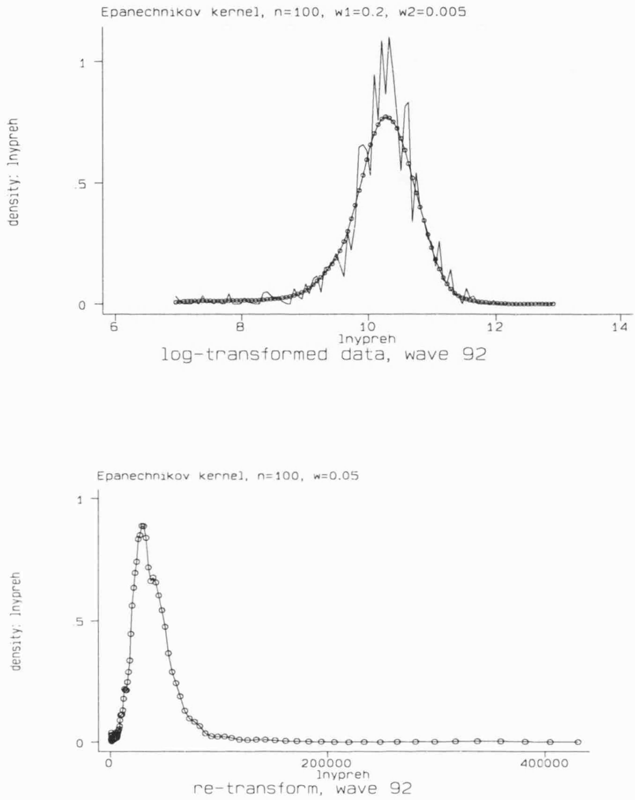


Figure 6.3: Upper panel: Density estimate of log-transformed net income with widths w.

Lower panel: Density estimate of net income with increasing bandwidth.

	year	1991	s
statistic	variable		
Gini	net equivalised income	.296	12.93*
	gross equivalised income	.334	
GE_{-1}	net equivalised income	.326	4.98*
	gross equivalised income	.533	
GE_0	net equivalised income	.168	4.05*
	gross equivalised income	.188	
GE_1	net equivalised income	.146	20.31*
	gross equivalised income	.202	
standard deviation	net equivalised income	19,007	
	gross equivalised income	33,928	
mean	net equivalised income	31,950	
	gross equivalised income	50,267	
median	net equivalised income	29,042	
	gross equivalised income	44,800	

Table 6.1: Summary statistics. s refers to the difference-of-means test (see chapter 4), * denotes statistically significant at the 5 percent level.

ing. This dominance is supported statistically using the IU inference rule. Recall from chapter 4 that s_i refers to the difference-of-means test of the Lorenz ordinates for population shares $i, i = 1, .., k$. All statistics s_i are positive and $\min\{s_1, .., s_k\}$ dramatically exceeds the critical level of the one-sided test with α=0.05 of 1.65.

Another welfare property can be investigated by inspecting the two empirical distribution functions (smoothed by joining the ordinates) depicted in Figure 6.4. The gross distribution nearly first-order stochastic dominates the net distribution. The only crossing happens at the very lowest income points, which is an unavoidable consequence of the fact that some households are entirely dependent on transfers. Again this difference is confirmed to be statistically significant by using a non-parametric two-sided Kolmogorov-Smirnov test.[3]

---

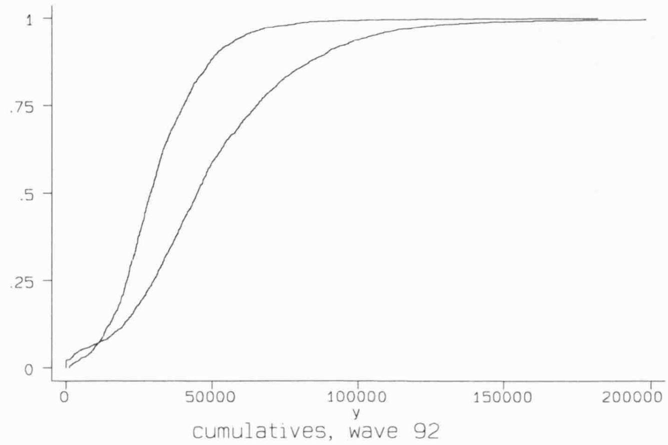
[3]The test statistic is

$$D := (0.5n)^{0.5} \max\{|\max_x\{F(x) - G(x)\}|, |\min_x\{F(x) - G(x)|\}$$

for distributions F and G and a sample of size n. Under the null hypothesis $F = G$, $\lim_{n\to\infty} \Pr\{D \leq \lambda_\alpha\}$. But the value D=2.4 exceeds even λ_α=2.2 at α=0.0001.

population share	net income Lorenz ordinate (SE)	income share (SE)	gross income Lorenz ordinate (SE)	income share (SE)	s
1	.029 (.0006)	.029 (.0006)	.0157 (.0007)	.0157 (.0007)	1377.7*
2	.083 (.0009)	.053 (.0005)	.0625 (.0011)	.0468 (.0005)	1307.3*
3	.149 (.0015)	.0663 (.0004)	.124 (.0013)	.061 (.0004)	1347.9*
4	.225 (.0015)	.0765 (.0004)	.197 (.0016)	.073 (.0004)	1329.2*
5	.312 (.0018)	.0867 (.0004)	.281 (.0018)	.084 (.0064)	1254.8*
6	.408 (.002)	.0965 (.0005)	.377 (.002)	.096 (.0004)	1133.6*
7	.517 (.0023)	.108 (.005)	.487 (.0022)	.111 (.0006)	937.1*
8	.641 (.0025)	.125 (.0006)	.616 (.0024)	.129 (.0006)	730.38*
9	.787 (.0027)	.46 (.0007)	.77 (.0024)	.154 (.0007)	469.2*

Table 6.2: Lorenz curve estimates 1991. Notes: Standard Errors in parenthesis, s refers to the difference-of-means test for each Lorenz ordinate (see chapter 4). * denotes statistical significance at the 5 percent level. The combined test evaluates to c=271.05*.



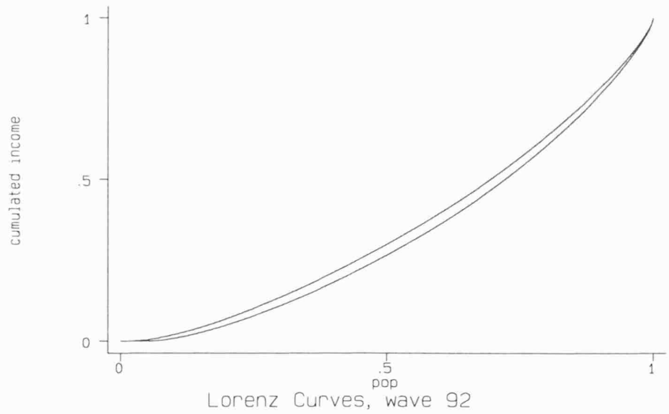


Figure 6.4: Upper diagram: The cumulative gross income distribution lies (almost completely) to the right of the net income distribution.

Lower diagram: The net distribution Lorenz dominates the gross distribution.

6.4 Explaining income inequality: The anatomy of inequality in 1991

The net distribution is likely to be a mixture of different underlying distributions as economic theory predicts that several characteristics have important distributional consequences. Since different individuals are at different points in their life-cycle, we expect age to play an important role. Human capital and discrimination theories highlight the importance of occupational status and nationality. Receipt of social security and welfare benefits is governed by eligibility conditions which only certain groups of non-earners can meet. Benefit levels differ also in terms of their generosity - most social security benefits are earnings related whilst means-tested social assistance benefits are flat rates. In order to investigate to what extent this population heterogeneity is important, the sample is partitioned accordingly.

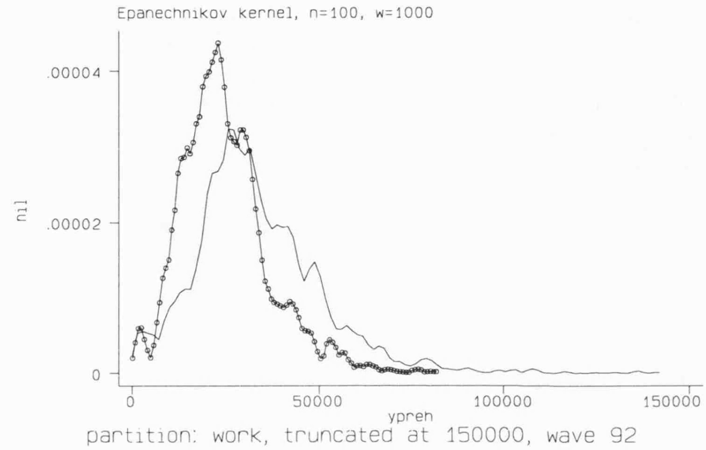
In Figure 6.5 the sample has been partitioned according to whether the household includes at least one earner. Not surprisingly, the density for earners is more dispersed than that for non-earners. The latter has a large mode about its median. This subsample falls into two categories, viz. households on means-tested social assistance in receipt of flat rate benefits and households receiving earnings-related social security benefits (i.e. unemployment benefit and pensions characterised by high replacement rates).

The sample depicted in Figure 6.5 is partitioned according to occupational status (which also proxies professional qualification or schooling because of the traditional strong link between the two and the highly segmented German labour market[4]). The density for blue collar workers has one dominant mode, whilst the density for white collar workers is multimodal with the modes being relatively far apart. This is not surprising given that this occupational category is spanned by six ascending levels of seniority.

In Figure 6.6, the sample has been partitioned according to whether or not the household contains at least one foreign national. As this figure makes clear, these households, the majority of which are classified as guestworker (having been recruited in the booms of the 1970s as manual workers), dominate the lower income groups

The population is partitioned in Figure 6.7 according to age. The cut-off point is 65 years, being the age of regular retirement. The income distribution of the elderly has a dominant mode about median income, which reflects the fact that the social security pension pay-

[4]As regards further education and other training programmes, Pischke (1994) examines both its incidence and duration using the special GSOEP module on training conducted in 1989. He finds that 61% of university graduates participate in training programmes, compared to only 22% of those with apprenticeships. Also, the duration of the programmes is highly skewed. On the cost side, 80% of all training is sponsored by employers and most training takes place during work hours. This training's profile potentially re-inforces wage and thus income inequality. However, this consequence apparently fails to materialise. "At least at this coarse level ... there is no relationship between training and productivity" (p.9) and Pischke concludes that "there is little evidence for wage gains among trainees" (p.3).



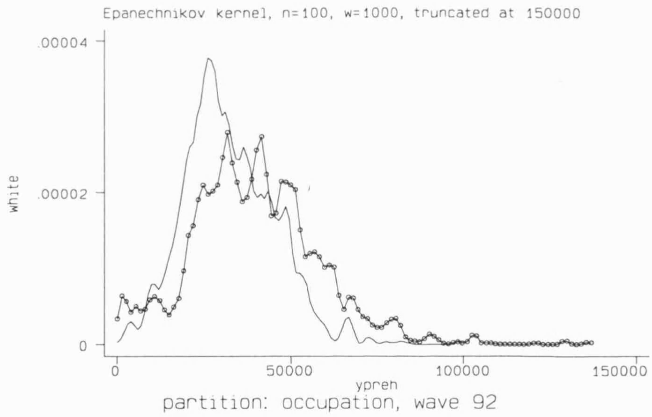


Figure 6.5: Density estimation of net income. Notes: circles in upper panel refer to households without earners; in the lower panel cirlcles refer to white collar workers.

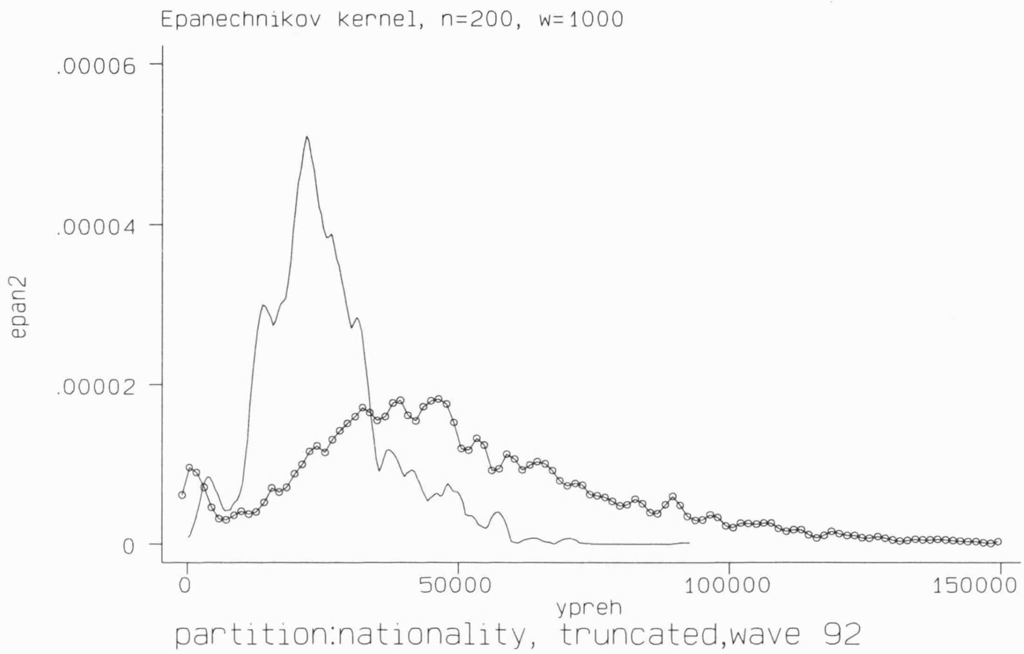


Figure 6.6: Density estimation of net income for Germans (circles) and foreign nationals.

ment, being earnings-related but with a ceiling on insured income, cannot exceed a certain amount. In the year 1991, this amount was about 4000 DM per month. Some people receive private pension payments which results in the upper tail. The distribution for the elderly is less dispersed than the income distribution of the younger.

Since age is a continuous variable, it seems natural to compute conditional inequality indices in order to capture other life-cycle characteristics [5]. Figure 6.8 depicts the generalised entropy index with several sensitivity parameters. All indices exhibit a similar pattern. Inequality increases in the forties but falls slightly in the mid fifties, as expected because seniority and human capital levels of earners

[5]The conditional statistics can be estimated using density techniques since $\hat{f}(y|a) = \hat{f}(y,a)/\hat{f}(a)$. Let h_y and h_a denote the bandwidths for the estimates of the respective marginal distributions. An estimator of the joint density is

$$\hat{f}(y,a) = (nh_y h_a)^{-1} \sum_i K(\frac{y - Y_i}{h_y}) K(\frac{a - A_i}{h_a}).$$

One can then solve for the conditional moments involved in the formula of the inequality index. For instance, $G_2(a) = 0.5(\mu_2(a)/\mu(a)^2 - 1)$. Using a Gaussian kernel, the conditional mean $\mu(a)$ is estimated by the Nadaraya-Watson estimator obtained by solving $\int y\hat{f}(y|a)dy$ (which requires a change of variable)

$$\hat{\mu}(a) = \frac{\sum_i K(\frac{a-A_i}{h_a})Y_i}{\sum_i K(\frac{a-A_i}{h_a})}$$

and the second conditional moment $\mu_2(a)$ by its obvious generalisation. More complicated cases can be solve by numeric integration. See Pudney (1992) for an extensive formal analysis of the life-cycle of inequality in China. Surprisingly, the shapes of his estimates are similar to the results depicted in Figure 6.8.

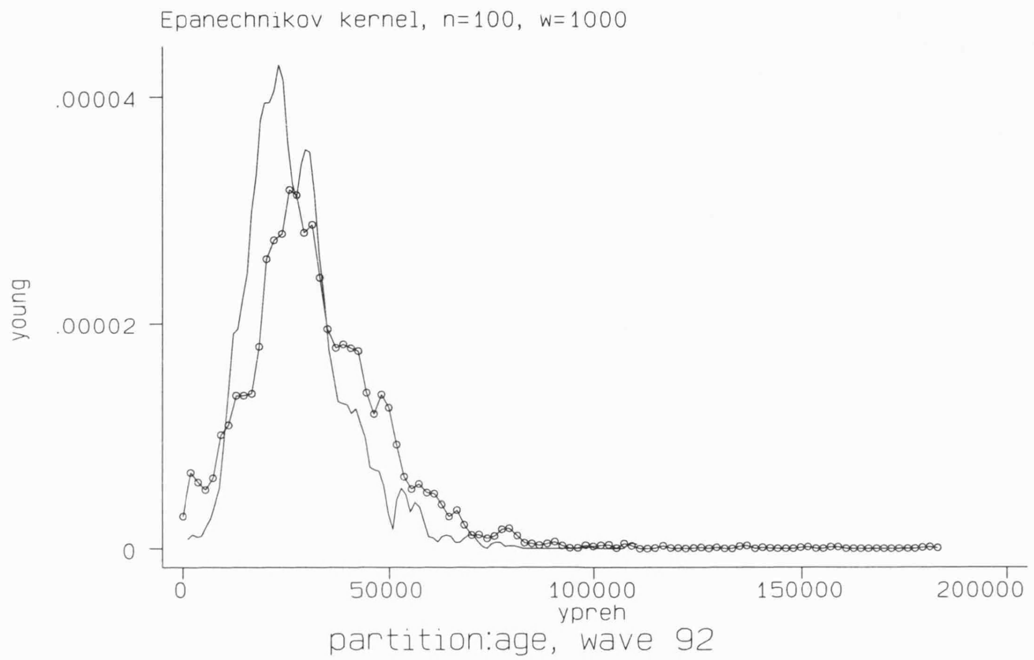


Figure 6.7: Density estimates for the elderly and the non-elderly (circles).

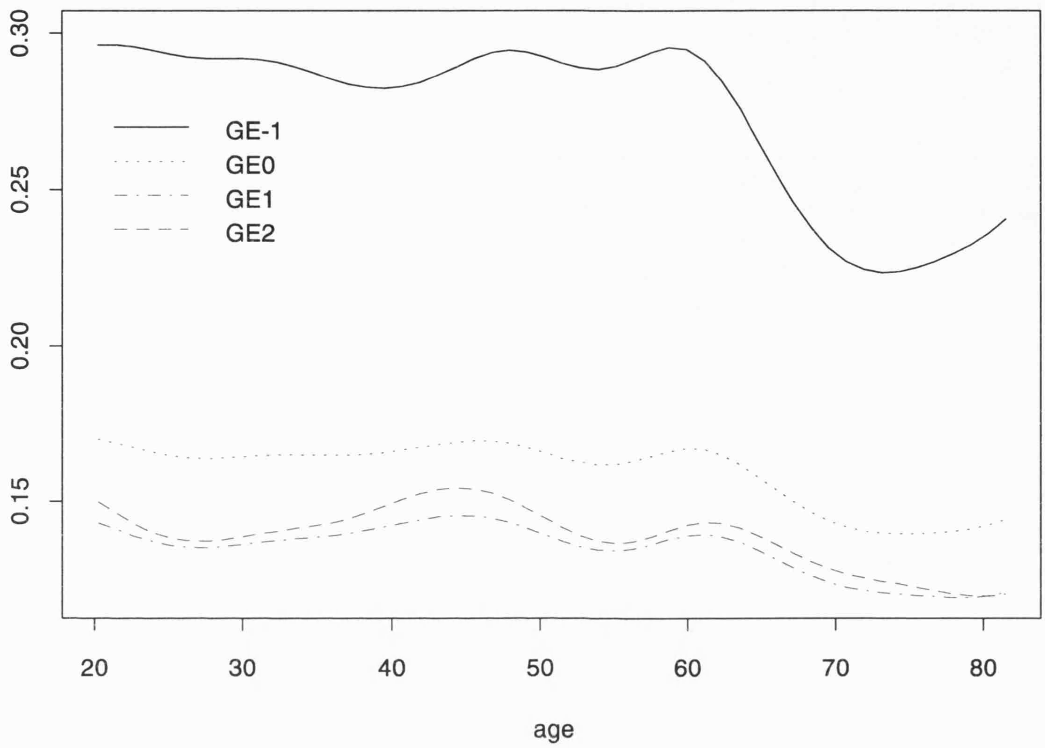


Figure 6.8: The generalised entropy index with several sensitivity parameter, conditional on age.

become more dispersed. A mitigating factor for this increased earnings inequality may be events within the household: children arrive in the family and jump into higher age brackets (which are taken into account by the McClements' scales). Around the age of retirement, inequality falls significantly as the elderly start to claim their pensions (although public pensions are earnings-related, previous earnings are insured only up to a ceiling).

A complementary analysis is to decompose the summary statistics of inequality according to the above partitions. Using the generalised entropy GE_{α} , overall inequality disaggregates into inequality within groups i, GE_{α}^{i}, and inequality between groups (when each group member has assigned mean group income) GE_{α}^{B} thus:

$$GE_{\alpha} = GE_{\alpha}^{within} + GE_{\alpha}^{between} = \sum_{i} p_i^{1-\alpha} y_i^{1-\alpha} GE_{\alpha}^{i} + GE_{\alpha}^{B}$$

where p_i represent the population share of group i, and y_i its income share. Table 6.3 reports the findings.

The separate point estimates GE_{α}^{i} register, independent of the sensitivity parameter α, a lesser degree of inequality within the groups of the elderly, households without work, and foreign nationals. However, comparing the within and between components of the inequality index reveals that overall inequality is driven by the within-component, as the young, employed, and German nationals have a more dispersed and unequal income distribution.

6.4.1 The German income distribution in context

The overall distribution of net income is a mixture of various underlying distributions. The low income groups are particularly dominated by foreign nationals, a sizeable group of the elderly, and those out of work. However, these forces tending towards greater inequality appear to be mitigated by the tax-benefit system, resulting in in-

partition	pop. share	income share	GE^i_{-1}	GE^B_{-1}	GE^i_0	GE^B_0	GE^i_1	GE^B_1
1. age ≤ 65 age > 65	.832 .168	.856 .1144	.372 .118	.002	.18 .094	.002	.153 .09	.002
2. 1^+worker no worker	.772 .229	.823 .177	.343 .234	.009	.168 .131	.007	.144 .109	.008
3. Germans foreigners	.935 .06	.948 .05	.335 .187	.002	.17 .117	.002	.147 .099	.002

Table 6.3: Decomposition of the generalised entropy index

equality indices for the entire distribution which appear to be low internationally. To back up this claim it is reasonable to compare the situation in Germany with the Uk and the USA.

In contrast to the unimodal German income distribution, the US and the UK income distributions have become bimodal in the 1980s. Burkhauser, Crews, Daly, and Jenkins (1995), Jenkins (1995), and Cowell, Jenkins, and Litchfield (1996) all find evidence of a shrinking middle class which is explained by an increase in the number of social assistance recipients and a concomitant increase in earnings inequality. So far, Germany has avoided this experience. Similarly, in a comparative study of relative trends of inequality using the Gini coefficient, Atkinson (1996) finds that German inequality has remained very stable in the 1980s. Inequality in the US has risen gradually. But the UK stands out for the sharpness of the recorded (monotonic) rise in inequality.

At a lower level of aggregation, we can compare the shape of the several income distributions depicted in Figures 6.5 and 6.6 with evidence for the UK reported in Cowell, Jenkins, and Litchfield (1996). In contrast to our with-without earner partition, they focus more narrowly on means-tested income support. As a consequence, the income distribution for this group is unimodal and very concentrated relative to the distribution for non-recipients. Our with-earner group is composed of individuals who are entitled to a variety of benefits. As all German social security insurance are earnings related, the relatively larger spread is not surprising. The conclusion from the kernel density estimates and the decomposition exercise supports the common view about inequality within the group of households without earners. Atkinson (1993a), however, shows that this common view is wrong in the case of the UK when all households out of work are considered (instead of focusing only on social assistance recipients). Although the UK awards mostly flat rate benefits, he demonstrates that the income distribution of those out of work has a long upper tail, so that it is deemed more unequal than the income distribution of those in work.

The comparison between the UK and Germany for the groups of the elderly and the non-elderly produces similar results. In the UK, the income distribution of the elderly is unimodal whereas the distribution for the non-elderly is bimodal and has a heavy tail. The mode for the elderly income distribution is more than twice as high as the one for the non-elderly. The shapes for the German case do not show such dramatic differences. The principal explanation for these different shapes is, again, the fact that the UK benefit system pays flat rate pensions (and the differences amongst the elderly arises from the receipt of private pensions), whereas German pensions are earnings related.

Consequently, this feature of the German tax-benefit system - the payment of earnings-related benefits- combined with the fact that earnings inequality has not increased substantially, explains why Germany has succeeded in producing relatively low and stable levels of inequality. By contrast, incomes in the US and the UK have become polarised. The main driving force behind this increasing inequality seems to be the increased number of recipients of flat rate social assistance benefits and an ever increasing earnings inequality.

6.5 Conclusions

The chapter sought to accomplish two tasks. First, the distributional consequences of the German tax-benefit system have been examined for the year 1991. The relevant distributions were estimated directly using a kernel density estimator and the net distribution was successfully tested for unimodality.

In an attempt to examine the welfare properties of the income distribution, we found that the net income distribution (statistically) Lorenz dominates the gross income distribution, so that all conventional inequality indices register a lower inequality of the net income distribution. The tax-benefit system thus constitutes an equalising force The gross distribution nearly first-order stochastically dominated the net distribution.

In the second step, the anatomy of the distribution of disposable incomes has been analysed. Several decomposition analyses show that the overall distribution of net income is a mixture of different underlying distributions. The low income groups are particularly dominated by foreign nationals, a sizeable group of the elderly, and those out of work. However, these coarse partitions reveal the considerable degree of inequality within the groups of the young, households with work, and German nationals, which dominates the assessment of overall inequality.

Against this background, chapter 7 analyses the issues of income dynamics by distinguishing between its two dimensions, viz. shape dynamics and intra-distributional mobility. Proceeding from this basis chapter 8 examines in detail income mobility and estimates models which accommodate time-varying transition probabilities and which might explain the observed mobility patterns.

Table 6.4:
inequality results in context

study	country	year	data source	income definition	equivalence scales	results
Frick et al.(1995)	West Germany	91	Gsoep	equivalised monthly net income + 1/12 of bonuses and perks	social assistance	Gini=0.263
		92				Gini=0.264
		93				Gini= 0.261
Buhmann et al.(1989)	West Germany	81	LIS/ Germany transfer survey	post-tax, post-benefit income	no	Gini=0.280
Jenkins (1994)	UK	90/91	HBAI	equivalised household net income	McClements	Gini=0.338 GE_0=0.227 GE_1=0.206
Buhmann et al.(1989)	UK	79	LIS/ FES	post-tax, post-benefit income	no	Gini=0.303
Atkinson (1993)	UK	88	FES	disposable income		Gini=0.350
	Finland	85	Household budget survey	disposable income	yes	Gini=0.200

For the comparison of the results, be it internationally or nationally, one important proviso needs to be taken into account, viz. sample selection and the treatment of the tails of the distribution. This is important, since most inequality indices fail to be robust against outliers (see Cowell and Victoria-Feser(1995)).

Chapter 7

Income Dynamics in Germany

Abstract: Using the GSOEP we analyse the two dimensions of German income dynamics in the 1990s by distinguishing between the shape dynamics and intra-distributional mobility. The complementarity between various analytical tools is explored. Although there appears to be an unambiguous increase in inequality, it is not statistically significant. By contrast, mobility has fallen significantly.

7.1 Introduction

A useful taxonomy for the analysis of distributional dynamics is the explicit distinction between shape dynamics - referring to the changing external shape of the distribution - and intra-distributional mobility. These two dimensions are orthogonal, require different types of data, and correspond to different sets of economic questions. An examination of the shape dynamics is a purely cross-sectional exercise, which implies that the set of addressable economic issues is quite limited (such as income polarisation, cross-sectional inequality or the incidence of poverty).

However, the increasing availability of longitudinal datasets has given a strong impetus to policy-related research based on the analysis of income histories. For instance, the extent of intra-distributional mobility is important for the design of welfare programmes. As most canonical models of the income or earnings process such as permanent income and life-cycle models are indeed dynamic, so should welfare assessments be. Lifetime equity depends on the extent of movement up or down the distribution. The above taxonomy can be fruitfully applied here because the analysis of the shape dynamics is inadequate for the problem of (lifetime) welfare assessments. This purely cross-sectional analysis cannot distinguish between such diametrically opposed worlds in which income positions are retained or permuted. It would rank the static economy exhibiting perfect persistence to possess the same level of welfare as the very mobile economy. But such a ranking would not conform to the value judgment of most people. Friedman (1962), for instance, considers "two societies that have the same distribution of annual income. In one there is great mobility and change so that the position of particular families in the income hierarchy varies widely from year to year. In the other, there is great rigidity so that each family stays in the same position year after year. Clearly, in any meaningful sense, the second would be the more unequal society" (p.171). A society may be better equipped to deal with short term fluctuations and a high degree of mobility - for instance consumption may be smoothed if credit and insurance markets are perfect- than with long term poverty.

Income mobility in Germany is low according to conventional wisdom. Such a prejudice often rests on the assumption that the labour market is segmented, inflexible and immobile; since earnings constitute the main source of income for the majority of the population, incomes should behave similarly. Moreover, the entire structure of the German welfare state is built on the premiss of stability - stable jobs and a consequent low income mobility. But does such a view fit the facts ?

In this chapter bivariate kernel density estimators are used to investigate directly the changes in the economic fortunes of individuals. But although such a tool constitutes a powerful device, it does not permit a rigorous statistical analysis. This is carried out after information has been aggregated in transition matrices. Changes in mobility are assessed using mobility indices.

This twin-track approach to income mobility is important since transition matrices are based on an arbitrary choice of income groups. Changing the partition of the sample by increasing the number of income groups or the length of the income intervals affects the transition matrix and consequently the mobility index employed in assessing intra-distributional mobility. By contrast, a stochastic kernel density estimate is more general since it does not impose any arbitrary grouping of information. Turning this argument around, these estimates should be used to give an indication of the extent to which the quantification of mobility based on transition matrices is robust to choices of income groups.

The plan of the chapter is as follows. For a description of the data we refer back to chapter 5. The movements of the cross-sections -the shape dynamics of the income distribution- are examined in section 7.2. The income distributions are estimated directly using (univariate) kernel density methods and inequality indices are computed in order to assess their welfare properties. In section 7.3 the twin-track approach to mobility is pursued. Section 4 concludes.

7.2 Cross sectional dynamics: a changing profile of inequality in the 90s?

[1]Has the shape of the net income distribution changed ? Figure 7.1[2] depicts the kernel density estimates of this distribution, which are superimposed on each other and thus permit a straightforward assessment of the shape dynamics of the income distribution. The shape of the distribution is surprisingly stable in the period 1990 to 1991, but changes slightly in 1992. Although the location of the (single) mode remains unchanged, more probability mass can be found to its left in the last year. The popular spectre of income polarisation - as expressed in catch words such as a 1/3 - 2/3-society - cannot be detected up to 1991 since the estimated distributions are unimodal [3], but a second mode in the lower income group emerges in 1992. However, these changes in the shape of the distribution are minor when compared to the vast changes experienced in the UK (Cowell, Jenkins, and Litchfield (1996)) or the US (Burkhauser, Crews, Daly, and Jenkins (1995)) in the 1980s. In both cases, a

[1]Chapter 8 contains some evidence for the period 1983 to 1989 for a slightly different income variable.

[2]The slight difference in the density estimates of the 1991 net income distribution in Figures 7.1 and 6.2 is due to the choice of different bandwidths. The latter density appears to be slightly under-smoothed.

[3]Chapter 6 confirms this unimodality for 1991 by applying Silverman (1981)'s bootstrap test for multi-modality. The second mode in 1992 is statistically significant.

dramatic polarisation has taken place; nearly unimodal shapes have turned into twin-peaks as the middle class occupies a sinking valley between them. As Jenkins (1995) observes: "The shift away from the middle class in both directions is strong evidence that the 'middle class' was shrunk, however one defines the middle." Fears of a shrinking German middle class are currently unfounded.

Table 7.1 [4] reports the key statistics [5] for the cross-sections in the years 1990 to 1992. GE_α refers to the generalised entropy index with sensitivity parameter α. There has been income growth -both mean and median income have risen- but what happened to income inequality ? Inequality appears to have unambiguously increased from 1991 to 1992, but the ranking of 1990 and 1991 seems to be ambiguous. Closer inspection of the values suggests that this ambiguity may just be caused by rounding errors and examining the respective Lorenz curves, reported in Table 7.2, supports this suspicion: the Lorenz curves do not intersect and shift out, but the ordinates for successive years are very close together.

At first glance it appears that the growth in income has been

[4]Frick and others (1995) produce some Ginis for this period (0.267, 0.263, 0.264) using a different income concept and equivalence scales based on the German social assistance programme. The story, however, is similar.

[5]The formulae for these measures are: Gini $=(1/(2n^2\hat{\mu}))\sum\sum|y_i - y_j|$ and $GE_\alpha = (1/(\alpha^2-\alpha))\left[(1/n)\sum(y_i/\hat{\mu})^\alpha - 1\right]$. If $\alpha = 0$ then $GE_0 = (-1/n)\sum\ln(y_i/\hat{\mu})$ as can be seen by using the transformation $\exp\{\ln(y_i/\hat{\mu})^\alpha\} = \exp\{\alpha\ln(y_i/\hat{\mu})\}$ and de l'Hospital's rule. Similar reasoning for α=1 yields $GE_1 = (-1/n)\sum(y_i/\hat{\mu})\ln(y_i/\hat{\mu})$.

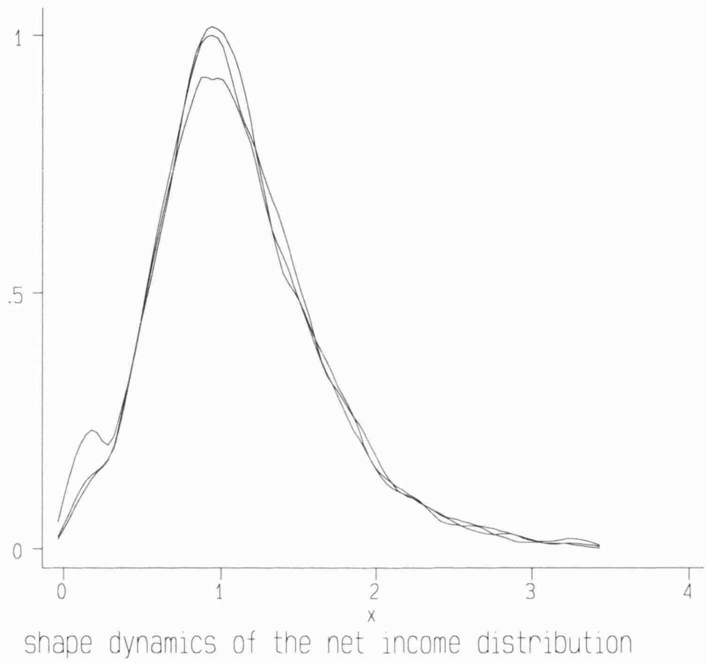


Figure 7.1: Notes: net income is normalised at the contemporaneous mean. The kernel density estimator uses the Epanechnikov kernel and the bandwidth is chosen using cross-validation methods.

year	1990	1991	1992	difference-of-means test		
statistic				$\lvert s_{90,91}\rvert$	$\lvert s_{90,92}\rvert$	$\lvert s_{91,92}\rvert$
Gini	.287	.286	.296	2.72*	2.2*	.48
GE_{-1}	.282	.287	.317	.68	1.65	.99
GE_0	.164	.168	.174	.26	1.13	1.39
GE_1	.148	.146	.157	.275	2.16*	2.42*
standard deviation	18,865	19,007	22,002			
mean	30,947	31,950	33,953			
median	28,125	29,042	30,650			
Note: * denotes significant at 5% level						

Table 7.1: Summary statistics and the pair-wise difference-of-means test

accompanied by a growth in inequality, as in many other OECD countries (Atkinson, Rainwater, and Smeeding (1994)), but are these changes in the inequality indices statistically significant? Since all the standard inequality indices are asymptotically normally distributed (see chapter 4), a non-parametric difference-of-means test can be applied to test this hypothesis. Table 7.1 reports the results of the pair-wise tests for the inequality indices. Most values are not statistically significant, or do not dramatically exceed the critical level of 1.96. Similarly, the (individual and combined) difference-of-means tests applied to the Lorenz ordinates reveals that the apparent differences are not statistically significant.

In summary, the apparent increase in inequality is not statistically significant. Does this lack of action on the surface - seemingly supporting the conventional wisdom of German immobility - conceal important intra-distributional changes?

popula tion share	year 1990 Lorenz ordinate	year 1991 Lorenz ordinate	year 1992 Lorenz ordinate	differenc e $s_{90,91}$	of means $s_{90,92}$	test $s_{91,92}$
.1	.031 (.0006)	.029 (.0006)	.024 (.0007)	2.32*	6.73*	5.2*
.2	.085 (.0009)	.082 (.0009)	.076 (.001)	1.64	4.06*	3.74*
.3	.152 (.0012)	.149 (.0012)	.14 (.0013)	.86	3.53*	3.5*
.4	.229 (.0015)	.225 (.0015)	.216 (.0016)	1.21	2.49*	1.89
.5	.316 (.0017)	.312 (.0018)	.301 (.0019)	-.88	0.72	1.04
.6	.412 (.0019)	.408 (.0020)	.396 (.0021)	.71	0.61	-.84
.7	.522 (.0022)	.517 (.0023)	.506 (.0024)	1.6	0.77	-1.3
.8	.646 (.0024)	.641 (.0025)	.632 (.0027)	-1.02	-2.0*	1.1
.9	.789 (.0026)	.787 (.0027)	.78 (.0029)	-2.24	-4.4*	1.79
c				5.5	20.1	3.2

Table 2: Lorenz curves and a pair-wise difference-of-means test.
Note: Standard errors are given in parenthesis. * denotes significantly different from zero on basis of a two-tailed test at 5% level. See the chapter 4 for a definition of the test statistics c and s.

7.3 Intra-distributional mobility

The second dimension of the distributional dynamics is intra-distributional mobility (Quah (1995)). It is also of particular importance for welfare assessments, since it is a measure of the inequality of opportunity and, according to some commentators, fairness. Basing welfare assessments solely on a cross-sectional analysis is inadequate since this method cannot discriminate between such diametrically opposed worlds where income positions are just replicated or, on the other hand, permuted.

The purpose of this section is to examine the folklore of German income immobility -whether German stasis is pervasive or whether inaction on the surface is deceptive. To this end, we first analyse a scatter plot depicting a typical example of the raw data. Then, a bivariate kernel density estimate depicts the movement of the entire distribution. Finally the analysis is supplemented at a higher level of aggregation by transition matrices and various mobility indices are computed in order to quantify the mobility process.

Transition matrices are a useful complement to kernel density estimates. On the one hand, transition matrices are needed for a statistical analysis of mobility changes and mobility indices are conventional tools for quantifying them. On the other hand, in transition matrices people are grouped according to their incomes, and within each group they are treated as if they were homogeneous. Both number and coverage of the income bands are imposed arbitrarily, and choosing different income bands results in different transition matrices. By contrast, the stochastic kernel density estimate (for each abscissa, there is a proper density which integrates to one) does not depend on such arbitrary grouping. A contour plot is depicted below in Figure 7.5 for two consecutive periods. Thus this twin-track approach to mobility is important as the strength of one technique compensates the weakness of the other.

Figure 7.2 maps the evolution of real equivalised disposable income from 1990 to 1991 as a scatter plot. A first ad hoc quantification of income mobility is to count the income transitions which exceed a critical level. To facilitate the inspection of the diagram, the 45 degree line has been superimposed. The second band represents a ±15% change in income which captures only 55.4% of the sample. A 10% and a 5% band capture only 42.8% and 25% respectively. An absolute band of ±5,000 DM (about 1/6 of median income) about the 45 degree captures 60% of the sample since a lot of income changes happen at the lower part of the joint distribution. Two results emerge:

First, the majority of incomes are persistent, falling into the ranges of the superimposed band but 75% of the population experience income changes in excess of 5%. This impression is corroborated when the bivariate distributions are estimated directly [6]. Figure 7.3

[6]The bivariate kernel density estimate was implemented following Silverman

superimposes the estimates for the years 1990-1991 and 1991-1992, whilst the lower panel of that figure furnishes a contour plot of the former. The contours appear to be symmetric relative to the 45 degree line, and the greatest concentration is found about median income in both periods. The distribution is unimodal and seems to have changed only slightly.

The plots of the kernel density estimates are supplemented by conventional transition matrices [7] reported in Table 7.3. Income has been grouped into four classes, the boundaries of which are defined in relative terms as 0.5, 1, 1.5 times the contemporaneous median income. This partition is useful since the poverty line is often defined in the applied literature in relative terms as 0.5 of median income. Moreover, the median is robust against outliers in the data.

(1986, p. 89) using the kernel

$$K(x) = \begin{cases} 3\pi^{-1}(1 - x^T x)^2 & if \quad x^T x < 1 \\ 0 & otherwise \end{cases}$$

Incomes are normalised at the contemporaneous median. The grid is 25x25 and the bandwidth is 0.7.

[7]These transition matrices are implemented using frequency counts. If the stochastic process is a non-stationary (first order) Markov process, the frequency count is also the maximum likelihood estimator (MLE) of the transition probabilities. Let $[p_{ik}(t)]$ denote the transition probability within an MxM transition matrix, and $n_{ik}(t)$ the number of persons moving from state i to state k at time t. The kernel of the likelihood function is $L = \prod_t \prod_i \prod_k p_{ik}(t)^{n_{ik}(t)}$, which is to be maximised subject to the constraint of $[p_{ik}(t)]$ being a stochastic matrix. This step yields the MLE $\hat{p}_{ik}(t) = n_{ik}(t)/\sum_k n_{ik}(t)$. Anderson and Goodman(1957) derive its asymptotic properties.

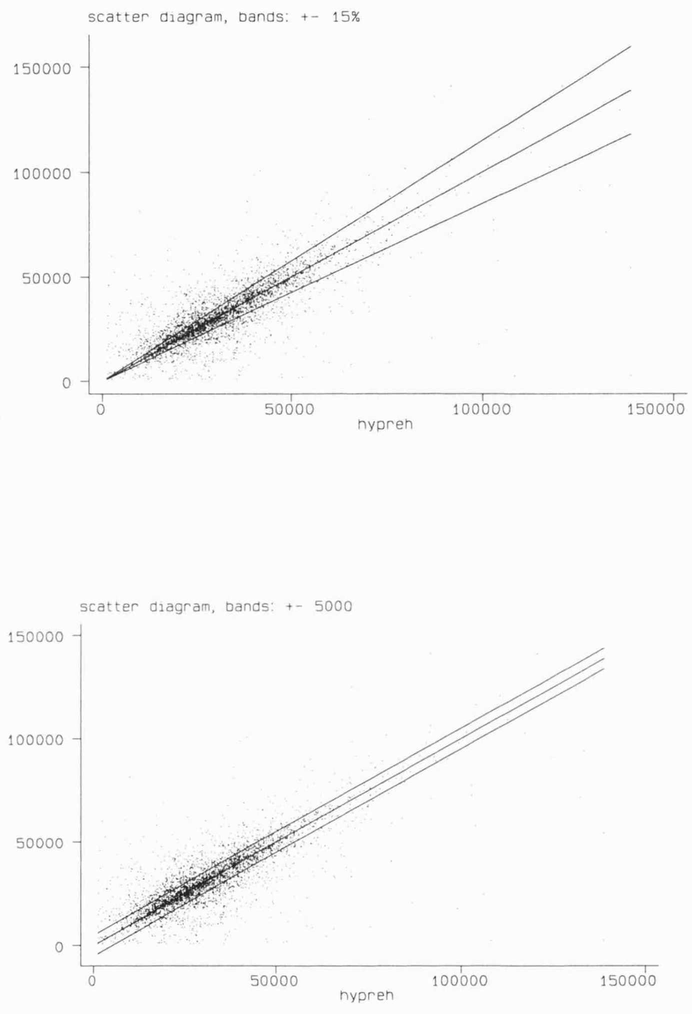


Figure 7.2: Scatter diagram for income transitions for year 1990 and 1992; Panel A: a band of relative deviations($\pm$ 15%); Panel B: absolute deviations ($\pm$ DM 5,000)

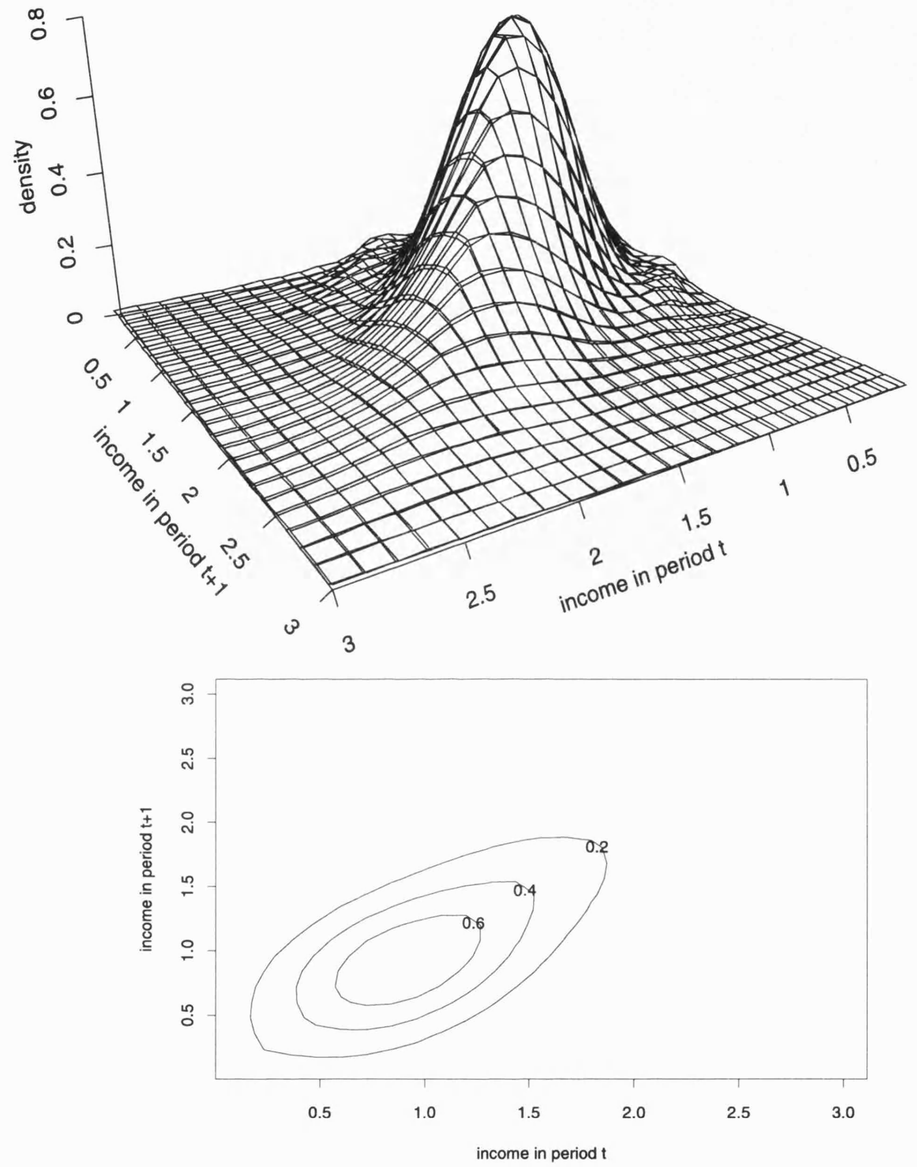


Figure 7.3: Upper panel: density estimates for 1990-1991 and 1991-1992. Lower panel: contour plot of the 1990-1991 density. Income is normalised at the contemporaneous median.

joint net income distribution 1990 and 1991

	year 1991				
year 1990	Low Income	Modest Income	Middle Income	High Income	Marginals
Low Income	.548	.329	.099	.023	.1
Modest Income	.107	.724	.154	.014	.39
Middle Income	.042	.246	.606	.011	.32
High Income	.021	.035	.217	.728	.18

joint net income distribution 1991 and 1992

	year 1992				
year 1991	Low Income	Modest Income	Middle Income	High Income	Marginals
Low Income	.549	.31	.11	.03	.12
Modes Income	.103	.79	.163	.016	.38
Middle Income	.044	.168	.667	.121	.31
High Income	.031	.04	.174	.754	.19

Notes: The four income groups are defined with respect to the contemporaneous median as follows:
low income (equivalised income below 0.5 times median income)
modest income (equivalised income between 0.5 and 1 times median income)
middle income (equivalised income between 1 and 1.5 times median income)
high income (equivalised income above 1.5 times median income)
weighted data

Table 7.3: Transition matrices

The diagonal elements of the transition matrices affirm the results of the contour plot: the underlying stochastic process shows both the strong broad persistence but also the significant risk of a change in one's income position. Persons in the lowest income group have a greater chance of escaping poverty, with a conditional probability of about $(1\text{-}p_{11})=0.45$. The other elements on the main diagonal exhibit strong persistence with a conditional probability of remaining in one's income group of about 0.7. The contour plot brings out more clearly the extent to which upward mobility is greater than downward mobility for the two poorest income groups but this is reversed for the two richest groups. These observations can be formalised by means of a mobility index. However, as with inequality indices, no single index is likely to be considered as completely satisfactory [8]. Let $P = [p_{ij}]$ denote the nxn transition matrix, λ_j its j's ordered eigenvalue, and $\mathrm{tr}(P)$ the trace of P. Shorrocks' index is defined as

[8] As Shorrocks(1978) observes:"(..) we may finally have to admit that no single mobility statistic has the minimum requirements regarded as essential"(p.1023). He shows that there is an inherent inconsistency between the notions of monotonicity [if $p_{ij} \geq p'_{ij}$ for all $i \neq j$ and with strict inequality for some, then the index should preserve this ranking] and perfect mobility [the index should have the maximum value for matrices with identical rows] as can be seen immediately by considering the 2x2 matrices p_{ij}=0.5 ,all i, j and p'_{ij}=0 ,$i = j$. In order to obtain a complete ordering he proposes to restrict the domain of admissable matrices to those possessing a "quasi-maximal diagonal". Moreover, as with inequality indices, all mobility measures incorporate value judgements. See Quah(1995) for a recent discussion of some mobility indices and some applications.

$$\mu_1(P) = \frac{n - tr(P)}{n-1} = \frac{n}{n-1}\left[\frac{1}{n}\sum_i (1 - p_{ii})\right] = \frac{n - \sum_i \lambda_i}{n-1} \quad (7.1)$$

being the inverse of the harmonic mean of the expected durations of remaining in a given income group. An alternative index is given by

$$\mu_2(P) = \frac{n - \sum_i |\lambda_i|}{n-1} \quad (7.2)$$

which equals μ_1 if P's eigenvalues are all real and non-negative. This index captures the speed of convergence of the underlying Markov process since all eigenvalues of P being a stochastic matrix, are bounded by one. This approach can be simplified by concentrating on the dominant convergence term, viz. the second largest eigenvalue λ_2

$$\mu_3(P) = 1 - |\lambda_2|. \quad (7.3)$$

This index would be attractive if the economy followed a (first order) Markov process. However, as it is demonstrated below, the transition probabilities are time-varying. Evaluating these indices yields $\mu_1(P_{90,91})= 0.43$ and $\mu_1(P_{91,92})= 0.41$, so mobility is deemed to have slightly fallen. This conclusion is inevitable since all staying probabilities p_{ii} have increased. The eigenvalues are respectively $\lambda_{90,91}$=(1; 0.75; 0.53; 0.42) and $\lambda_{91,92}$=(1; 0.76; 0.54; 0.48). Since all elements are real and non-negative μ_1 and μ_2 are identical. The use of μ_3 results in the same ranking of the transition matrices.

In order to assess the statistical significance of the changes confidence intervals are estimated by bootstrapping and the extent of their overlapping is examined. Using 1000 bootstrap replications Efron's (1987) bias-corrected-and-accelerated BC_α method [9] is ap-

[9]See Efron(1987) and Efron and Tibshirani (1993) for a description of this method. The confidence interval derived by this method has two principal characteristics: it is invariant to transformations, and its error decays at rate 1/n.

The following paragraph gives an outline of this method. B=1000 bootstrap replications -a random draw with replacement- are drawn from the original income pairs, and the transition matrix and μ_1 is computed for each draw. Let $\hat{\mu}^*$ denote the bootstrap sample estimator of the mobility index for the population parameter μ, and $\hat{\mu}$ the value of the estimator using the original data. Then the BC_α method provides a $(1-2\alpha)$ confidence interval $(\hat{\mu}^{*(\alpha1)}; \hat{\mu}^{*(\alpha2)})$, where $\hat{\mu}^{*(\alpha)}$ denotes the αth-percentile of the bootstrap sample. α_1 and α_2 are determined as follows. Let Φ denote the cumulative normal distribution, Φ^{-1} its inverse, and $z(\alpha)$ its 100 αth percentile point. Then

$$\alpha_1 = \Phi\left(\hat{z} + \frac{\hat{z} + z^{(\alpha)}}{1 - \hat{a}\left[\hat{z} + z^{(\alpha)}\right]}\right) \quad ; \quad \alpha_1 = \Phi\left(\hat{z} + \frac{\hat{z} + z^{(\alpha)}}{1 - \hat{a}\left[\hat{z} + z^{(\alpha)}\right]}\right)$$

How are $\hat{z}$ and $\hat{a}$ determined ? $\hat{z}$ is given by $\hat{z} = \Phi^{-1}((\#(\ \hat{\mu}^* < \hat{\mu})/B))$. $\hat{a}$, however, is given in terms of the jacknife values -resampling without replacement- of the statistic. Let $x_{(i)}$ denote the sample derived from the original sample but having the i-th value removed, let $\hat{\mu}_{(i)}$ be calculated using $x_{(i)}$, and average $\hat{\mu}_{(.)} = (1/n)\sum \hat{\mu}_{(i)}$. Then $\hat{a}$ is given by

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\mu}_{(.)} - \hat{\mu}_{(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\mu}_{(.)} - \hat{\mu}_{(i)}\right)^2\right]^{3/2}}$$

$\hat{z}$ is a measure of the median bias of $\hat{\mu}^*$, and $\hat{a}$ deals with the rate of change of $\hat{\mu}$ with respect to the population parameter μ.

plied to the Shorrocks' index μ_1 (but could equally well be applied to μ_2). The resulting 90% confidence intervals are [0.41; 0.45] for $\mu_1(P_{90,91})$ and [0.3856;0.4228] for $\mu_1(P_{91,92})$, which overlap. Nevertheless, the difference of means test, described in the appendix, suggests that the estimates of the index are statistically significantly different.

Another way of assessing the changes in the mobility structure is to superimpose the stochastic kernel density estimates. Figure 7.5 depicts the contour plots for 1990-1991 and 1991-1992. There are large changes in the probability of staying in poverty which tend to increase. There is also a remarkable stability for income groups II and III. Inference for changes above 2 can be drawn because of the small numbers of observations. In general, different groups are affected differently, since there is no universal trend (which underscores the need to work with general tools such as stochastic kernels instead of relying solely on transition matrices subject to arbitrary grouping).

This brief discussion leads naturally to the next issues about the structure of the transition matrices: the two transition matrices appear to be roughly similar. A multinomial test [10] could be applied in order to render precise the extent to which these transition matrices

[10]See, for instance, Mood et al.(1974, p.449). It is desired to test the null hypothesis $p_{1j} = p_{2j} = p_j$ for two populations $i = 1, 2$, with associated probabilities p_{ij}, $j = 1, .., k+1$. Sample i has size n_i and N_{ij} denotes the number of outcomes in group j. p_j needs to be estimated; its maximum likelihood estimator is $(N_{1j} + N_{2j})/(n_1 + n_2)$. The test statistics can then be arrived at as

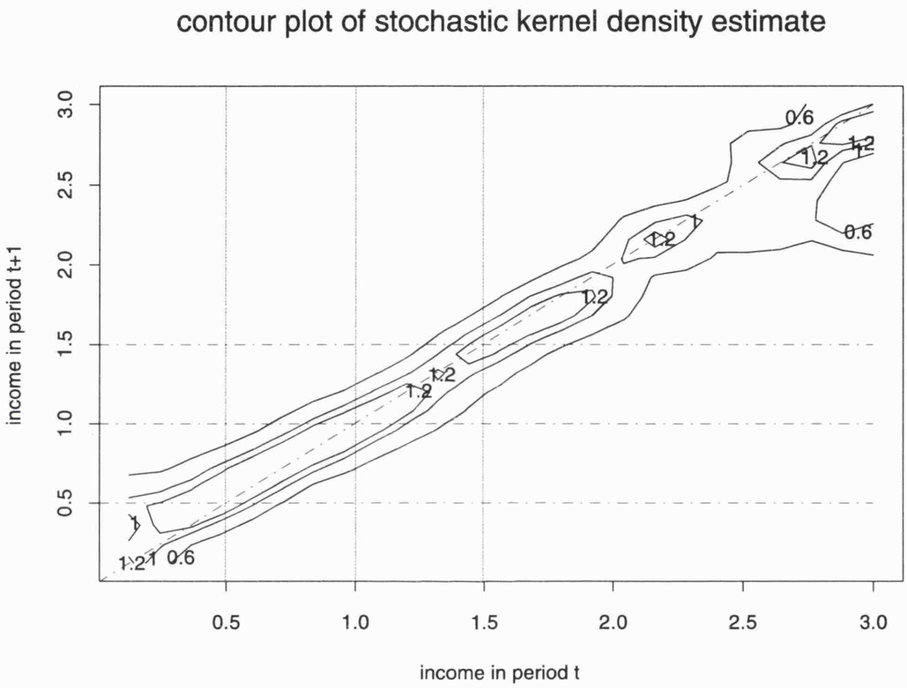


Figure 7.4: The contourplot of a stochastic kernel density estimate for 1990-1991 ; the superimposed grid represents the income groups of the transition matrices.

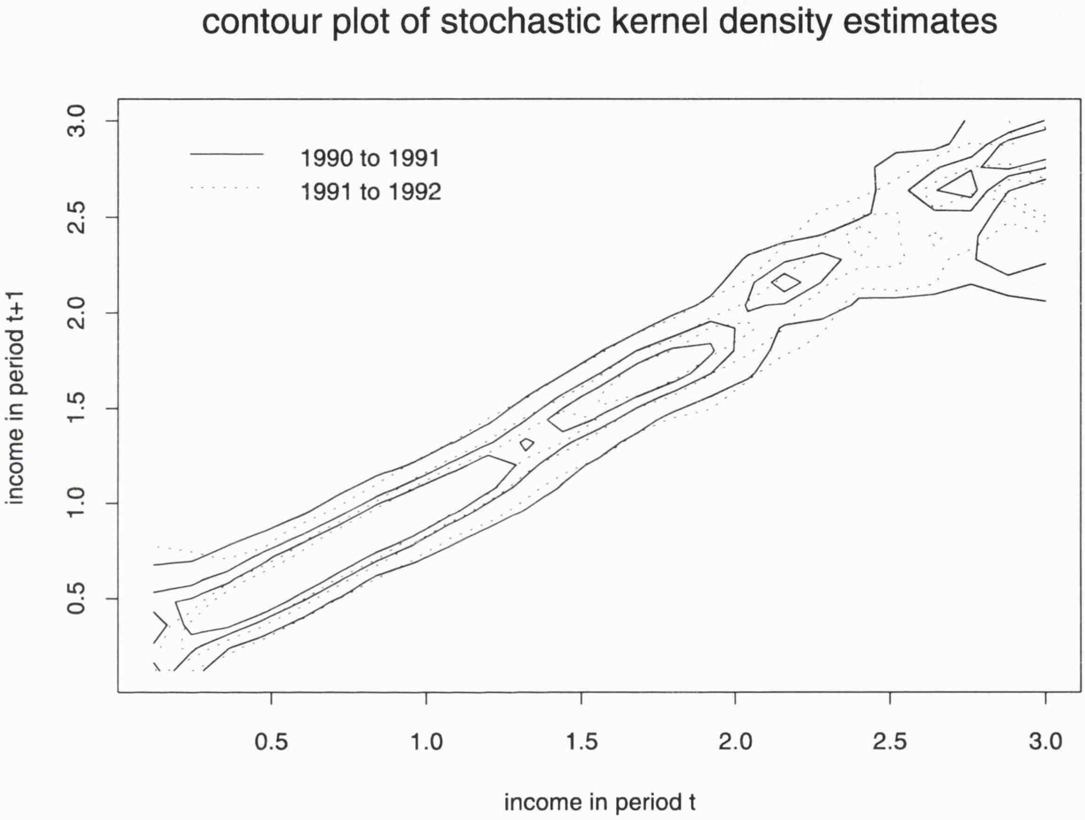


Figure 7.5: Contour plot of the stochastic kernel density estimates for income transitions 1990-91 and 1991-92.

are identical. The values of the test statistics for the separate tests are 4.01, 17.56*, 19.51*, 6.75, and for the combined test 47.82*. The three starred items are significantly different from zero, leading to the conclusion that the two matrices are not identical.

Are the transition matrices symmetric, i.e. $p_{ij} = p_{ji}$? Inspection suggests otherwise, since there is evidence of a dominant upward mobility for the first two income groups: The values of the likelihood ratio test statistics [11] are 62.65* and 31.9* respectively, exceeding the critical value of 16.8 at the 1% level.

These tests suggest that the transition matrices are not symmetric and that the transition probabilities are time-varying. These sta-

---

$$S_l = \sum_{i=1}^{2}\sum_{j=1}^{k+1} \frac{[N_{ij} - n_i(N_{1j} + N_{2j})/(n_1 + n_2)]^2}{n_i(N_{1j} + N_{2j})/(n_1 + n_2)}$$

which has a chi-square limiting distribution with k degrees of freedom. The test can be combined into one (see, for instance, Amemiya (1985, p.417)). The null hypothesis is that both matrices are identical. The test statistics is $\sum_l S_l$ for an MxM transition matrix, the summation ranging over the number of rows, which is distributed as chi-squared with $M(M-1)$ degrees of freedom. An alternative test is a likelihood-ratio test, see Anderson and Goodman (1957).

[11] See, for instance, Bishop, Fienberg, and Holland (1988). The test is of the form

$$X^2 = \sum_{i>j} \frac{(x_{ij} - x_{ji})^2}{x_{ij} + x_{ji}}$$

where x_{ij} represents an observation on a multinomial variate. The test statistic is distributed as chi-squared with $M(M-1)/2$ degrees of freedom, M being the size of the transition matrix.

tistical results are important and suggest that the process governing income dynamics is rich and complicated.

German stability? A reassessment

Contrary to the conventional wisdom, there is much mobility beneath the seemingly stable surface of the German income distribution. This is good news for a (dynamic) welfare assessment of the income distribution.

But the bad news is the distributional dynamics in the 1990s. Income inequality rises whilst mobility falls. This seems to be a general trend amongst the OECD countries (Atkinson, Rainwater, and Smeeding (1994)). For instance, "income mobility decreased substantially" (Veum (1992)) in the US. The driving force behind the income changes appears to be changes in earnings and Buchinsky and Hunt (1995) uncover "a sharp decrease in mobility over time, across all skill groups" (regardless of educational attainment and experience). A similar conclusion is reached by Gosling, Machin, and Meghir (1996) for the case of the UK who conclude that the increased wage inequality -risen "at an unprecedented scale"- can only be explained by an increase in the relative importance of the unobservable component determining wages. As in the US, changing returns to education or skill cannot explain the extent of the observed dispersion. These detrimental effects have been amplified by a reduction in the generosity of welfare benefits -the second principal component of income- and the onset of the economic downturns (1991 in the

German case) Other important features are summarised in Atkinson, Bourguignon, and Morrisson (1992). These distributional trends should be of concern to policy makers.

7.4 Conclusions

This chapter has analysed the two dimensions of income dynamics by distinguishing between the shape dynamics and intra-distributional mobility. The analysis of the moving cross-section reveals that income inequality has not significantly changed although the point estimates appeared to suggest an unambiguous increase in inequality as the Lorenz curves shifted out. The density estimates for 1990 and 1991 are unimodal, but in 1992 a second mode emerges in the low income group.

Pursuing a twin-track approach to mobility - stochastic kernel density estimates and transition matrices - we showed that Germany is more mobile than conventional wisdom suggests. This is good news, but the bad news is that mobility is falling. The bivariate density estimates are unimodal. A person's income position is persistent because both contour plots of bivariate kernel density estimates and transition matrices concentrate most mass along the main diagonal. On the other hand 75% of the population experienced income changes in excess of 5% - a fact which is in stark contrast to the popular prejudice of German income immobility. As regards the structure of the transition matrices, they are neither symmetric nor identical as transition probabilities vary with time. This non-stationarity is confirmed by standard mobility indices, which suggest that mobility has unambiguously fallen. The useful complementarity between transition matrices and stochastic kernel density estimates were explored, since the latter do not impose arbitrary groupings of information. The contour plots of the latter showed the complexity of the income process as different income groups are affected differently; at this level aggregation, there is no universal trend such as a greater concentration about the 45 degree line.

Recalling Friedman's criterion for welfare assessments quoted in the introduction, we must conclude that Germany has become a "more unequal society".

Against this background, chapter 8 examines in detail the issue of non-stationarity in income transitions, and analyses to what extent popular stochastic models - such as higher Markovian or mover-stayer models - can account for this. In an attempt to explain the observed mobility patterns, various duration models are examined.

Chapter 8

On the Non-Stationarity of German Income Mobility (and some observations on poverty dynamics)

Abstract: The intra-distributional mobility of German income dynamics is analysed using GSOEP. Transition probabilities are found to be time-varying. The tested models comprise various mixed Markov chains in discrete time and a non-stationary mover-stayer model is proposed. In order to explain the observed mobility profiles, we concentrate on one important income class -the poor- instead of the entire transition matrix. Various poverty duration models are examined.

8.1 Introduction

As argued in the last previous chapter, the study of income dynamics can be divided into the distinct tasks of examining the shape dynamics of the income distribution, and investigating intra-distributional mobility. Whilst the former enjoys much popularity because cross-sectional data is often readily accessible, the latter task is more elusive, and a common prejudice is that (at least in Germany) mobility is low. The German Socio-Economic Panel (GSOEP) is used in this chapter to examine income mobility for the German case. This characterisation is made in two stages. The first stage is descriptive and follows an established literature which represents social processes by Markov models [1] (and a new mover-stayer model is proposed). The second stage goes beyond mere description and attempts to explain the observed mobility.

Some stylised facts about German income dynamics in the period 1983 to 1989 can be established by examining Figures 8.1 to 8.3. The shape of the net income distribution has hardly changed in this pe-

[1] See, for instance, Champernowne (1973), McCall (1971), the references in the eponymous paper by Singer and Spilerman (1974), Geweke, Marshall, and Zarkin (1986). Quah has applied the above taxonomy in a series of papers in the context of the debate about the (non)convergence amongst countries. See, for instance, Quah (1996).

riod [2]. But in contrast to this seeming 'stability', the time series of summary statistics of the transition matrices, such as the Shorrocks (1978) mobility index depicted in Figure 8.2, reveal that mobility dynamics behave differently. The lack of action at the surface conceals substantial movements beneath it. Several features emerge. The statistic is not a constant, suggesting that the underlying transition probabilities are time-varying. In fact, there is a downwards trend which implies a consistent fall of income mobility over the years except for the last year. Since the Shorrocks mobility index is the inverse of the harmonic mean of expected durations of remaining in a given part of the cross-section distribution, the lower panel of the figure depicts the time series of the staying probabilities. For the three richest income groups, these have a tendency to rise, but not monotonically and some movements are in opposite directions. By contrast, the lowest income group - the poor - experience a dramatic increase in immobility, but there is also a sharp fall in the last year.

Another way of assessing income mobility is to examine the con-

[2] The changes in the shape of the distribution are minor when compared to the vast changes experienced in the UK (Cowell, Jenkins, and Litchfield (1996)) or the US (Burkhauser, Crews, Daly, and Jenkins (1995)) in the 1980s. In both cases, a dramatic polarisation has taken place; nearly unimodal shapes have turned into twin-peaks as the middle class occupies a sinking valley between them. As Jenkins (1995) observes: "The shift away from the middle class in both directions is strong evidence that the 'middle class' was shrunk, however one defines the middle." Fears of a shrinking German middle class are currently unfounded.

tour plots of the stochastic kernel density estimates, depicted in Figure 8.3. They are more general than transition matrices as they do not impose an arbitrary grouping of information. There is a remarkable stability in the middle income group but important changes happen in the lower and upper income groups. All this evidence suggests that the mobility process is very complicated.

Such changes in the mobility dynamics are clearly important when welfare assessments of income distributions are not static, but attempt to capture income histories. Changing income mobility indicates changing opportunities, and according to some commentators changing fairness[3]. Concentrating exclusively on the shape dynamics of the income distribution - seemingly unchanged in the German case - would ignore one important dimension of welfare.

Several models are tested which might explain these stylised facts. Section 8.3 and 8.4 pursue a statistical approach and examine whether pure and mixed Markovian models in discrete time give an adequate representation of the mobility process. In order to account for population heterogeneity, a (new) non-stationary mover-stayer model. We pursue the twin-track approach to mobility which exploits the important complementarity between stochastic kernel density esti-

[3]Friedman (1962), for instance, considers "two societies that have the same distribution of annual income. In one there is great mobility and change so that the position of particular families in the income hierarchy varies widely from year to year. In the other, there is great rigidity so that each family stays in the same position year after year. Clearly, in any meaningful sense, the second would be the more unequal society" (p.171).

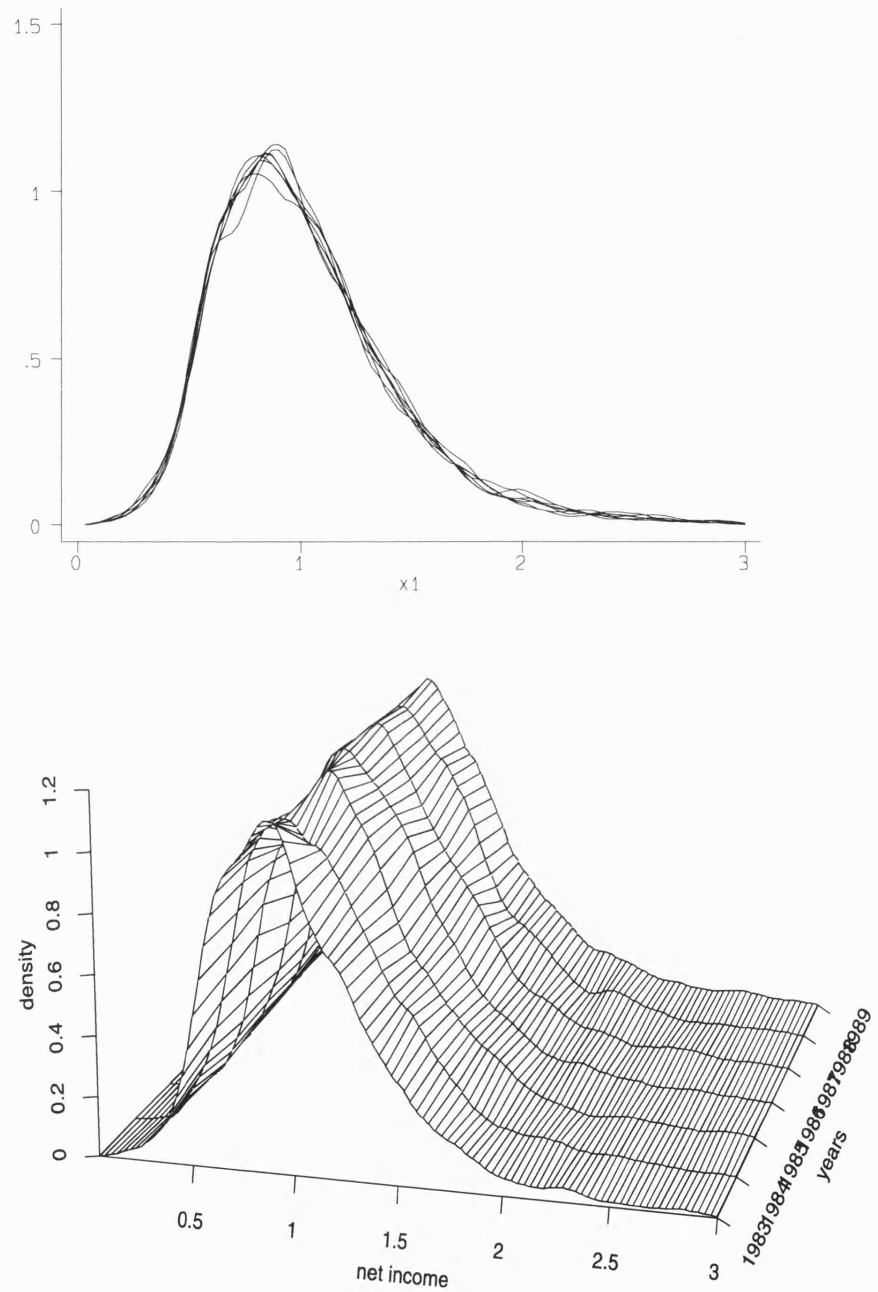


Figure 8.1: Shape dynamics of the net income distribution. Net income is normalised at the contemporaneous mean. The kernel density estimator uses the Epanechnikov kernel and the bandwidth is chosen using cross-validation methods.

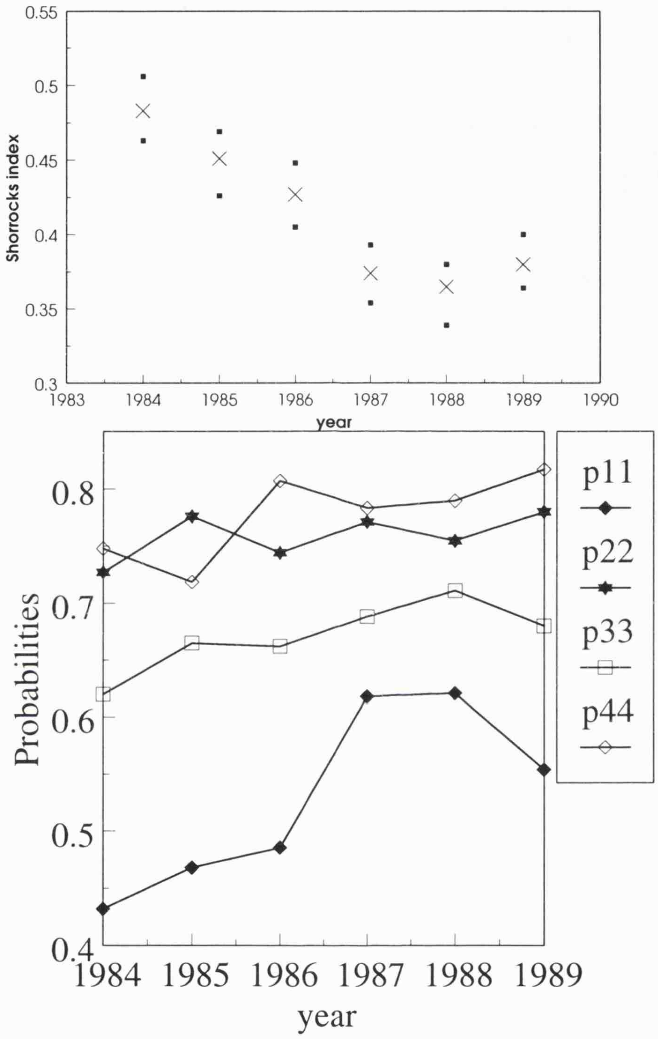


Figure 8.2: Upper Panel: Time series of the Shorrocks mobility index, bracketed by estimated 10% confidence. intervals. Lower Panel: Time series of staying probabilities.

Notes: If $P = [p_{ij}]$ denotes the nxn transition matrix, the Shorrocks index proposed in (Shorrocks 1978) is defined as $\mu(P) = (n - tr(P))/(n-1)$. It is the inverse of the harmonic mean of expected durations of remaining in a given part of the cross section distribution. The higher the index, the lower is the persistence or the greater is the mobility. The transition probabilities are estimated using their maximum likelihood estimator (cf. section 3). The confidence intervals were computed using bootstrapping and (Efron 1987)'s BC_α

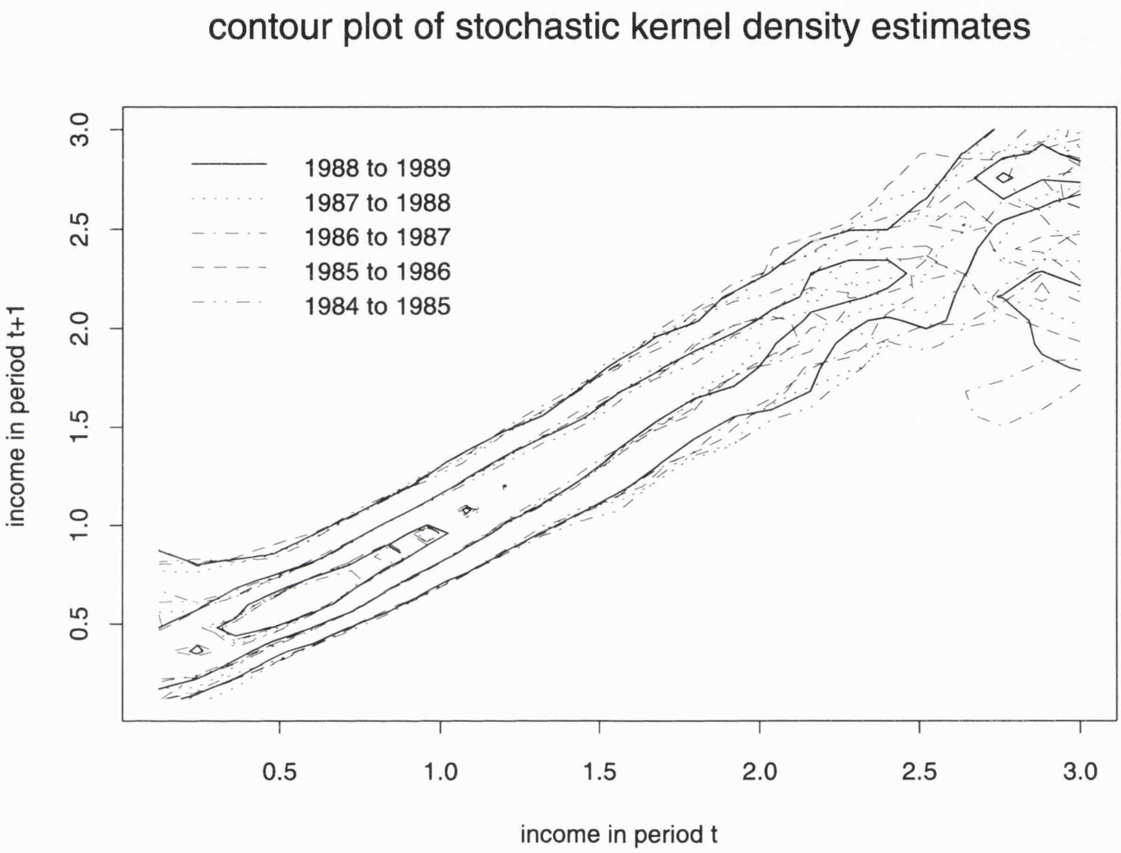


Figure 8.3: contour plots of the stochastic kernel density estimates 1984-1985 to 1988-1989. Income is normalised at the contemporaneous median.

mates and transition matrices as the strength of one tool compensates the weakness of the other. Although the mobility analysis is based on transition matrices, the income partition is not only chosen in an economically meaningful way, but also happens to coincide with the natural partition suggested by Figure 8.3. On the other hand, transition matrices are a powerful tool for making rigorous statistical inferences. The next section attempts to explain the observed mobility. The method pursued here is to concentrate on one important income state instead of the entire transition matrix, and the chosen income state is poverty -the income state about which most policy makers are concerned and for whose alleviation considerable resources are deployed. Persistent social exclusion is recognised as a grave problem facing any society. Moreover, this group has experienced the most dramatic changes in mobility. Section 8.5.1 explains the chances of escaping or descending into poverty by means of a Markov model which accounts both for the non-stationarity of the data and the heterogeneity of the population. A second class of models, analysed in section 8.5.2, comprises duration models. Section 8.6 concludes.

8.2 The data

In this chapter the German panel dataset GSOEP is used in its incarnation as the "Equivalent Datafile". Comprising the years 1984 to 1990, the latter is a subset of the former and not only contains its principal income variable but also includes some derived variables, the most important being post-tax post-benefit income. Since GSOEP proper was described in detail in chapter 5, a brief outline should suffice. Two income concepts are used. The elements of annual gross (pre-tax pre-benefit) income are raw data but need to be aggregated. However, the Equivalent Datafile conveniently supplies an estimate of annual household post-tax post-benefit income, which is derived from the gross income data by means of a tax-benefit simulation. This income variable is computed as the sum of total family income from earnings, asset flows, private and public transfers, the imputed rental value of owner occupied housing, and a tax simulation is applied. [4]

In order to take account of scale economies within the household, income was equivalised using the OECD equivalent scales. Disposable income was divided by household size raised to the power 0.5. This choice of equivalence scales had been made for two reasons. First, Burkhauser, Merz, and Smeeding (1994) show that the German Social Assistance scale implies scale economies which are too low. Second, the use of the OECD scale, being the standard scale for datasets included in the LIS project, facilitates first ad hoc international comparisons.

Finally, incomes were standardised at 1991 prices. The data remained unweighted for the subsequent estimation procedures. The

[4]The income variable differs from slightly the one used in chapter 6 and chapter 7 in that the former includes income from asset flows.

sample examined in this chapter was selected by keeping only persons with a complete income record for the years 1984 to 1990. This selection procedure resulted in 9022 observations. Subsequently, the income data is analysed by means of transition matrices. Four income groups were defined with respect to the contemporaneous median (a statistic which is robust against outliers). The poverty line is set (arbitrarily) at 0.5 times median income. Modest incomes are equivalised incomes between 0.5 and 1 times the median. Middle incomes are between 1 and 1.5 times the median. Finally, high incomes are those above 1.5 times the median. The choice of these income groups is inherently arbitrary, but the relative definition of poverty applied in this chapter has become standard practise for European countries. In fact, this partition is also suggested by Figure 8.3.

8.3 Pure Markov models in discrete time

This section explores the extent to which standard Markovian models can explain the observed income transitions. Competing models are juxtaposed, and the following sequence of tests is conducted: non-stationary and stationary Markov chains of the same order are tested against each other. Then, a non-stationary first order chain is tested against a non-stationary second order chain. These tests were first developed by Anderson and Goodman (1957).

8.3.1 First order Markov chains

Let $P(t) = [p_{ij}(t)]$ be an mxm transition matrix where $p_{ij}(t)$ denotes the conditional probability of moving to state j in the current period, given that state i was occupied in the preceding period. The chain is observed up to time T at time points $t = 1, 2, .., T$. If the chain is stationary $p_{ij}(t_1) = p_{ij}(t_2) = p_{ij}, \quad \forall t_1, t_2$. Let $N(t) = [n_{ij}(t)]$ denote the associated matrix of actual transition counts. The transition probabilities $p_{ij}(t)$ need to be estimated. Since these are multinomially distributed, their maximum likelihood estimator $\hat{p}_{ij}(t)$ can be derived by maximising the likelihood function, conditional on the initial distribution

$$\log L = \sum_t \sum_i \sum_j n_{ij}(t) \cdot \log p_{ij}(t)$$

subject to $P(t)$ being a stochastic matrix. The Lagrangian for this programme is

$$\mathcal{L} = \sum_t \sum_i \sum_j n_{ij}(t) \log p_{ij}(t) - \sum_t \sum_i \lambda_{ti} \left(\sum_j p_{ij}(t) - 1 \right)$$

The first order condition implies $n_{ij}(t) = \lambda_{ti} \cdot \hat{p}_{ij}(t)$. Summing out j yields $\lambda_{ti} = \sum_j n_{ij}$, which upon substitution gives the estimator

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{\sum_j n_{ij}(t)}$$

being a simple frequency count. For the stationary model, a similar calculation gives the maximum likelihood estimator [5]

$$\tilde{p}_{ij} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} \sum_{j} n_{ij}(t)} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} n_{i+}(t)}$$

where $\sum_j n_{ij}(t) = n_{i+}(t)$ for notational convenience.

Given these functions, tests for non-stationarity can be easily implemented. Let the null hypothesis be that the transition probabilities are stationary, i.e. $\mathcal{H}_0$: $p_{ij}(t) = p_{ij} \quad \forall t, i, j$. Using the respective likelihood functions, the likelihood ratio is

$$\log \lambda = \sum_{t} \sum_{i} \sum_{j} n_{ij}(t) \cdot \{\log \tilde{p}_{ij} - \log \hat{p}_{ij}(t)\}$$

and $-2 \log \lambda$ is asymptotically distributed as χ^2 with $(T-1)m(m-1)$ degrees of freedom[6]. If the null hypothesis is true, an asymptotically equivalent test is based on the similarity between transition matrices and contingency tables. The well-known χ^2-test then gives the test statistic

$$X^2 = \sum_{i} X_i^2 = \sum_{i} \sum_{t} \sum_{j} n_{i+} \cdot \frac{[\tilde{p}_{ij} - \hat{p}_{ij}(t)]^2}{\tilde{p}_{ij}}$$

which is asymptotically distributed as χ^2 with $m(m-1)(T-1)$ degrees of freedom. However, Anderson and Goodman (1957) show

[5] Anderson and Goodman (1957) show its asymptotic sampling distribution to be normal.

[6] See Rao (1973). The number of degrees of freedom of the asymptotic χ^2-distribution equals the number of linearly independent restrictions. Note that P is a stochastic matrix.

that if the null hypothesis is not true, the power of the χ^2 -test can be different from the power of the likelihood test. Thus both tests should be performed.

Since a direct inspection of the transition matrices or the time series of the mobility index suggests non-stationarity, it is not surprising that both tests confirm the greater explanatory power of the non-stationary model. The test statistics evaluate to $-2\log\lambda = 336.3$ and $X^2 = 341.4$.

8.3.2 Second order Markov chains

The methods of the preceding paragraphs extend in a very natural manner to second order Markov chains. $P(t) = [p_{ijk}(t)]$ denotes the conditional probability of being in state k at time t, given states i and j at times $t-2$ and $t-1$ respectively. Again, the transition probabilities are estimated by maximising the log-likelihood function, and similar likelihood ratio and χ^2- tests apply.

The tests suggest once again that the non-stationary (second order) model has a greater explanatory power than the stationary model. The test statistics evaluate to $-2\log\lambda = 561.4$ and $\chi^2 = 545.7$.

8.3.3 First order against second order Markov chains

The theoretical results of the two preceding sections can be combined in order to test which order of the non-stationary model has the greater explanatory power. If the null hypothesis is that a first order non-stationary model is applicable, $p_{1jk} = p_{2jk} = \ldots = p_{mjk} = p_{jk}, \forall j, k$, the likelihood ratio becomes

$$\log \lambda = \sum_t \sum_i \sum_j \sum_k n_{ijk}(t) \cdot \{\log \hat{p}_{jk} - \log \hat{p}_{ijk}\}$$

$-2 \log \lambda$ being asymptotically distributed as χ^2 with Tm^2 degrees of freedom.

Performing this test, the statistic evaluates to $-2 \log \lambda = 2,955.8$, being very significant evidence against the null hypothesis. In consequence, the memory of the process governing income transitions extends over more than one period.

8.4 Mixed Markovian models in discrete time

Pure Markovian models are popular both in the theoretical as well as in the empirical literature because of their mathematical structure. Yet, as the previous section demonstrated, they do not fit the data too well. One principal assumption underlying their estimation is the 'homogeneity of persons': individuals are the same except for their income. This assumption is likely to be flawed. Indeed, results derived in chapter 6 suggest that the population is very heterogeneous. In contrast to this observable heterogeneity, some latent variable may be important. A particular type of unobservable heterogeneity is treated next.

The next level of complexity is achieved by mixing independent Markovian models, the easiest of which is the following mover-stayer model. An unobservable fraction of the population stays with certainty in its income group for all periods, whilst the evolution of incomes of everyone else, the movers, is determined by a non-degenerate first order Markov chain. The pure Markovian model is nested within this richer structure, since stayers may not be present. This nesting gives rise to a natural test, a likelihood ratio test, by means of which to discriminate between these two models.

8.4.1 A mover-stayer model : the stationary case

Although Goodman (1961) presents an extensive mathematical treatment of this model, his estimators, proposed without derivation, are not maximum likelihood estimators. These estimators are supplied in Frydman (1984) where transitions are stationary.[7] A non-stationary

[7] McCall (1971) applies the mover-stayer model to the issue of earnings mobility. He simplifies the estimator proposed in Goodman (1961) by letting $T \to \infty$ despite the fact that his empirical time series is very short.

model is proposed below.

Let the unobserved fraction of the population who are stayers in income group i be denoted by s_i. The income of movers $(1 - s_i)$ evolves according to the stationary first order Markov chain with mxm transition matrix $M = [m_{ij}]$. The composite process thus evolves according to $P(t) = SI + (I - S)M^t$ where S is a diagonal matrix with entries s_i. Let $n_i(t)$ denote the number of persons in state i at time t, n_i the number of persons staying in state i during the entire period of observation, $n_{ik} = \sum_t n_{ik}(t)$ the total number of transitions from state i to state k, $n_i^* = \sum_t n_i(t-1)$, and n the total number of persons. The log-likelihood function conditional on the initial distribution can be factorised thus

$$\begin{aligned} \log L(s, M) &= \sum_i n_i \log\left[s_i + (1 - s_i) m_{ii}^T\right] \qquad (8.1) \\ &+ \sum_i \left[n_i(0) - n_i\right] \log(1 - s_i) + \left[n_{ii} - T n_i\right] \log m_{ii} \\ &+ \sum_i \sum_{k \neq i} n_{ik} \log m_{ik} \end{aligned}$$

The last summation pertains only to transition between unequal states, and thus concerns only movers. As regards the first sum, a person may remain in income class i for two reasons: either he is a stayer with probability s_i, or with probability $(1 - s_i)$ he is a mover but remains in that state for T consecutive periods with probability m_{ii}^T. The second term captures movers returning to their initial state, who have at least once left it.

The maximisation strategy is to resubstitute solutions from the first order conditions into the objective function[8]. Eventually the size of the equations system is reduced to the number of income classes, and the equations for $\hat{m}_{ii}$ can be solved numerically. The estimators of the off-diagonal elements $\hat{m}_{ij}$ are then computed recursively

$$\hat{m}_{ij} = n_{ij}(1 - \hat{m}_{ii} - \sum_{\substack{k=1 \\ k \neq i}}^{j-1} \hat{m}_{ik}) / \sum_{\substack{k=j \\ k \neq i}} n_{ik}$$

The estimates for the stayers are[9]

$$\hat{s}_i = \frac{n_i - n_i\,(0)\,\hat{m}_{ii}^T}{n_i\,(0)\,(1 - \hat{m}_{ii}^T)} = 1 - \frac{n_i\,(0) - n_i}{n_i\,(0)\,(1 - \hat{m}_{ii}^T)} \tag{8.2}$$

the last term being the ratio of the observed to the expected number of persons who make a least one transition from state i during T periods.

Since the pure Markovian model is nested within the mover-stayer model, their relative performance can be assessed using a likelihood ratio test. Let the null hypothesis be that the pure model is appropriate ($s_i = 0, \forall i$). The maximised likelihood of the pure model is

[8] Amemiya (1985) proposes an alternative method of estimation. Two equation systems are established by considering transitions within two periods, viz. $P = SI + (I - S)M$ and $P^{(2)} = SI + (I - S)M^2$. P and $P^{(2)}$ can be consistently estimated by the maximum likelihood estimators presented in the preceding sections. Resubstituting these yields $2m(m-1)$ equations in m^2 unknowns. However, Frydman's estimation strategy is more parsimonious.

[9] This estimator may become useless if the observation period T is small, since $\hat{s}_i$ may become negative as Frydman failed to impose a non-negativity constraint. This problem decays with m_{ii}^T as T increases.

$\log(L_{s=0}) = \sum_i n_i\,(0) \log\left(n_i\,(0) - n\right) + \sum_{i,k} n_{ik} \log\left(n_{ik}/n_i^*\right)$. Denote the ratio of the likelihoods by λ, $-2\log\lambda$ is distributed as χ^2 with m degrees of freedom.[10] Can the statistical significance of an individual s_i be tested ? If T and $\hat{m}_{ii}$ are such that $\hat{m}_{ii}^T$ is negligible, the estimator for s_i simplifies to $\hat{s}_{i=}n_i/n_i\,(0)$, being the fraction of persons initially in state i who remain there for all consecutive periods. In this case Goodman (1961)'s argument applies. Let $\tilde{p}_{ij}$ denote the estimator of the stationary first order Markov chain derived in the previous section. The test is based on a comparison between $\hat{s}$ and its expected value $\tilde{p}_{ii}^T$. $(\hat{s} - \tilde{p}_{ii})$ is normally distributed with mean zero and a variance which can be consistently estimated by $\hat{\sigma}^2 = \tilde{p}_{ii}^T(1-\tilde{p}_{ii}^T)/n_i\,(0) - n\tilde{p}_{ii}^{2T-1}(1-\tilde{p}_{ii}^T)/\bar{n}_i$, where $\bar{n}_i = \sum_t n_i\,(t)\,/T$. Under the null hypothesis $s_i = 0$, and $X_i^2 = (s - \tilde{p}_{ii}^T)/\hat{\sigma}^2$ is distributed as χ^2 with one degree of freedom.

The estimated transitions matrix is not reported here for the sake of brevity. Compared to the pure model, probability mass has been redistributed away from the main diagonal. The movers are thus more mobile than the pure model suggests. The stayers fractions are estimated to be $\hat{s} = (0.1; 0.224; 0.04; 0.14)$. Testing the competing models, the likelihood ratio test confirms the greater explanatory power of the mover-stayer model ($-2\log\lambda$ =23,147.8). These estimate have a profound implication, since the first income group is

[10]See Rao (1973). The number of degrees of freedom of the asymptotic χ^2-distribution equals the number of linearly independent restrictions. There are m restrictions imposed, viz. $\{s_i = 0\}_{i=1}^m$.

occupied by the poor (whose income falls short of the contemporaneous poverty line). Income mobility is sufficiently high so that most persons are able to escape poverty at least temporarily. Yet, a statistically significant 10%[11] of those deemed in poverty at the beginning of the observation period constitute a hard-core of poverty[12]-remaining poor with certainty.

These results, of course, have to be taken with a pinch of salt, as the previous section suggested that income transitions are non-stationary. The problem caused by time-varying transition probabilities is addressed in the next section.

8.4.2 A mover-stayer model: the non-stationary case

The previous model can be generalised so that non-stationarity in income transitions can be introduced. Let movers transit according to the non-stationary first order Markov chain $M(t)$. In consequence, the composite process evolves according to $P(t) = SI + (I - S)\prod_{\tau=1}^{t} M(\tau)$. Analogous to equation (8.1), the likelihood function can be written as

$$\log L(s, M(1), ..., M(T)) = \sum_i n_i \log\left(s_i + (1 - s_i)\prod_\tau m_{ii}(\tau)\right)$$

[11] Applying the above test to the estimate of s_1, the estimate of the hard-core of poverty, yields a statistically significant result ($\chi_1^2 = 3,766.8$).

[12] Labelled by McCall (1971) the "back-wash hypothesis".

$$+\sum_i (n_i(0) - n_i) \log (1 - s_i)$$
$$+\sum_i \sum_\tau (n_{ii}(\tau) - n_i) \log m_{ii}(\tau)$$
$$+\sum_i \sum_\tau \sum_{k \neq i} n_{ik}(\tau) \log m_{ik}(\tau) \quad (8.3)$$

Observe that the stationary case of equation (8.1) is nested within equation (8.3). Maximising this with respect to s_i yields the estimator

$$s_i = \frac{n_i - n_i(0) \prod_\tau m_{ii}(\tau)}{n_i(0)\left(1 - \prod_\tau m_{ii}(\tau)\right)} \quad (8.4)$$

(compare to (8.2).) Resubstituting this into equation (8.3) yields

$$\begin{aligned} \log L(\hat{s}, M(1), ..., M(T)) &= c - \sum_i (n_i(0) - n_i) \log \left(1 - \prod_\tau m_{ii}(\tau)\right) \\ &+ \sum_i \sum_\tau (n_{ii}(\tau) - n_i) \log m_{ii}(\tau) \\ &+ \sum_i \sum_\tau \sum_{k \neq i} n_{ik}(\tau) \log m_{ik}(\tau) \end{aligned}$$

where c denotes a constant. The Lagrangian of this problem is

$$\mathfrak{L} = \log L(\hat{s}, M(1), ..., M(T)) - \sum_i \sum_\tau \lambda_i(t) \left(\sum_k m_{ik}(t) - 1\right)$$

Maximising this with respect to $m_{ii}(t)$ and $m_{ik}(t)$, summing these and solving out the Lagrange multipliers $\lambda_i(t)$ yields a non-linear equations system for $m_{ii}(t)$

$$n_{ii}(t) - n_i = (n_{i+}(t) - n_i) m_{ii}(t) + \left[n_i(0) - n_i \frac{\prod_\tau m_{ii}(\tau)}{1 - \prod_\tau m_{ii}(\tau)}\right] m_{ii}(t)$$

This is a non-linear system comprising T equations in T unknowns with solution $\hat{m}_{ii} = (\hat{m}_{ii}(1), ..., \hat{m}_{ii}(T)) \in [0,1]^T$. It is solved numerically using the multidimensional Newton's method (see Press, Teukolsky, Vetterling, and Flannery (1992)). The solutions to $\hat{m}_{ik}(t)$ and $\hat{s}_i$ are then computed recursively.

Tests of hypotheses can be implemented following the methods outlined in the previous sections.

This estimation procedure resulted in the following estimates for the stayers $\hat{s} = (0.11; 0.33; 0.13; 0.31)$. Compared to the stationary model, the estimates of the fraction of the poor has not changed, whilst the other estimates all have increased. As regards the mover probabilities, the entries of $M(t)$ follow the changes suggested by the movement of the Shorrock mobility index. Probability mass is moved onto the main diagonal as time passes, suggesting that incomes have become more immobile.

8.5 Poverty re-examined

The aim of this section is to go beyond the descriptive Markov models of the preceding sections and to attempt to explain the observed mobility profiles. Instead of analysing the entire transition matrix, we concentrate on one important income state - poverty - and analyse the processes governing the movements into and out of poverty. Two types of models are examined.

8.5.1 A Markov model with observed heterogeneity

This section examines a two state Markov model with exogenous variables as proposed in Boskin and Nold (1975) and further discussed in Amemiya (1985). Person i may be in either of two states: either he is in poverty at time t, $y_i(t) = 1$, or he is not $y_i(t) = 0$. The probability of being in poverty conditional on the preceding state is $\Pr(y_i(t) = 1|y_i(t-1)) = F(\beta' x_i(t) + \gamma' x_i(t)\, y_i(t-1))$ where $F(.)$ is a distribution function with corresponding density f. Thus, the model is a generalised first order Markov model, in which the exogenous variables $x_i(t)$ exhibit non-stationarity and heterogeneity amongst persons. This formulation nests within it a variety of observationally equivalent models, depending on the parametrisation of γ. For instance, setting $\gamma = -(\alpha + \beta)$ and if f is symmetric, the model has the following interpretation. The (conditional) probability of person i entering poverty is determined as $p^i_{01}(t) = F(\beta' x_i(t))$, whereas the (conditional) probability of escaping poverty is $p^i_{10}(t) = F(\alpha' x_i(t))$. Thus the profile of a representative person entering poverty is stipulated to be different from that of a representative person escaping poverty.

The log-likelihood function can be written as

$$
\begin{aligned}
\log L(\alpha, \beta) \;=\; & \sum_i \sum_t y_i(t) \log F(\beta' x_i(t) + \gamma' x_i(t)\, y_i(t-1)) \\
& + (1 - y_i(t)) \log \left[1 - F(\beta' x_i(t) + \gamma' x_i(t)\, y_i(t-1))\right]
\end{aligned}
$$

The distribution function is chosen to be logistic $F(x) = e^x/(1 + e^x)$, so that the objective function is globally concave and the estimation step reduces to estimating a standard logit model. The maximisation strategy is to employ the iterative method of scoring separately for each parameter. The MLE is consistent and asymptotically normal (see Amemiya (1985)). Note also that the indices $(i; t)$ can be treated as a single index. Thus, although the time series is relatively short but the cross section is large, the sample can be considered to be large.

The sample was chosen to contain only persons above the age of 20 in order to focus on the causes of poverty, a step which reduces the size of the sample to 6266 observations. The regressors comprise: indicators for employment status, disability, and household size in a given year [13], nationality, the age, and education level (measured in years) of the person in the year 1984. The importance of these variables is not surprising given the results of a static analysis in Schluter (1996a) who estimates the income distributions for various partitions of the sample using kernel density estimators.

[13] Bane and Ellwood (1985), for instance, emphasise the importance of the household formation process as a determinant of poverty in the case of the US.

Table 1: Maximum likelihood estimates for poverty model

unweighted model	**variables (standard errors)**							**chi-squared**	**log-likelihood**	**N**
parameter	nationality	age84	hhsize	education84	disability	unemployment	constant			
β	.886 (.076)	-.027 (.002)	-.304 (.02)	-.131 (.02)	.437 (.08)	1.53 (.077)	-1.24 (.29)	716.48	-4,174.4	35,574
γ	.268 (.11)	.006 (.003)	-.137 (.036)	-.085 (.027)	.203 (.11)	.64 (.11)	.57 (.38)	114.3	-1,337.9	2,022
α=-(β+γ)	-1.154	-.033	.441	.216	-.64	-2.17	.67			

weighted model	**(robust standard errors)**									
β	.859 (.1)	-.038 (.003)	-.55 (.05)	-.14 (.027)	.363 (.11)	1.75 (.124)	-.034 (.418)		-4,340.1	35,574
γ	.43 (.156)	.008 (.004)	-.15 (.053)	-.103 (.033)	.08 (.15)	.727 (.169)	.75 (.45)		-1,302.5	2,022
α	-.1.289	.03	.697	.24	-.443	-2.47	-.72			

Notes: The variables are defined as follows: nationality is an indicator set equal to one for foreigners; age84 is the age of a person in year 1984 on the survey date; the education level is measured by years of education up to year 1984; hhsize refers to the size of the household at time t disability at time t is an indicator equal to one for disabled persons; unemployment at time t equals one when unemployed.

Chi-squared refers to the likelihood ratio test, the null hypothesis being that only the constant term has explanatory power.

Robust standard errors were computed following the methods proposed in (Huber 1967).

Both weighted and unweighted data are used and Table 8.1 collects the estimation results. The results contain some surprises. As regards the unweighted data, the probability of escaping poverty is higher when a person is a German, is well educated, healthy and ends unemployment spells quickly. This last ability is the most decisive and the relative size of the parameter estimate is perhaps astonishing. More formally, the relative importance of the variables can be assessed by computing an elasticity such as $\eta_{i;j} := (\partial p_{01}^{i}(t)/\partial x_{j}^{i}(t))/(x_{j}^{i}(t)/p_{01}^{i}(t)) = \beta_{j}(1 - p_{01}^{i}(t))x_{j}^{i}(t)$ which approximates the effect of a change in a discrete variable x_j for person i. The effect of becoming unemployed is dramatic: $\eta_{i;unemployed} = 1.53(1 - p_{01}^{i}(t))$ (but this effect diminishes as $\beta' x^{i}(t)$ increases). Unemployment and nationality are of even greater importance for those escaping poverty.

The coefficient for nationality is large but this may be due to oversampling foreigners. When the data is weighted, the coefficient on nationality is expected to fall because foreigners were oversampled. Surprisingly, the coefficient is only slightly lower, but the employment status coefficient is markedly higher. This implies a different profile for persons slipping into poverty. For Germans, the principal reason appears to be unemployment, whilst it seems to be low earnings for foreigners.

8.5.2 Semi-Markov processes: non-stationary duration models

The discrete time models of the preceding sections have to confronted the time aggregation problem, highlighted in Singer and Spilerman (1976), caused by the absence of a natural time unit: income transitions do not happen at the end of regularly spaced intervals which coincide with those of the panel survey. As a consequence, parameter estimates cannot be interpreted as structural information. In this case it is more appropriate to fit a continuous time model. But this strategy gives rise to two problems. First, the model needs to be formulated in such a way that the actual discrete time observation is embeddable within the continuous time model. A nice set of necessary and sufficient conditions has yet not been found. [14]

The second problem is caused by the particular data under scrutiny, viz. their non-stationarity. Whilst it is impracticable to estimate a general continuous time Markov chain, researchers have pursued two avenues. Singer and Spilerman (1976) discuss (but do not estimate) a mixture model in which transitions follow a stationary Markov chain but waiting times between transitions may vary with time. A second possibility and the strategy pursued below is to focus on one economically meaningful state, such as poverty, and to estimate a

[14]Geweke, Marshall, and Zarkin (1986), for instance, present a calculation to test the embeddability of a discrete first order stationary Markov chain within a stationary continuous time model. See also their references for the embeddability problem.

parametrised duration model.

This section presents a standard duration model as outlined in Cox and Oakes (1984) and follows suggestions of Amemiya (1985). Person i may be in either of two states: either he is poor or he is not. The time in poverty T, i.e. the length of the poverty spell, is a random variable with distribution F and associated density f. If the population is heterogeneous, these may differ across persons, written as F_i. It is convenient to work with the hazard rate $\lambda_i(t) := f_i(t) / [1 - F_i(t)]$ where $\lambda_i(t)\,\Delta t$ has a probabilistic interpretation: it is the probability that, given the person has not left poverty in the time interval$(0,t)$, he will do so the next moment, i.e. in $(t, t + \Delta t)$. A basic assumption of the continuous time model is reminiscent of Poisson processes, since the probability that a person changes her income state more than once in a small time interval $(t, t + \Delta t)$ is negligible. $\lambda_i(t)$ may vary with time. The duration function F can then be written as

$$F_i(t) = 1 - \exp\{-\int_0^t \lambda_i(z)\,dz\} \tag{8.5}$$

If person i completes J poverty spells of individual length $t_{i,j}$ the contribution to the likelihood function is $\prod_{j=1}^{J} f_i(t_{i,j})$. However, the estimation problem is complicated by the fact that person i may have censored spells. A spell at the end of the panel t^* is right-censored and thus incomplete if the person cannot be observed to leave that state, leading to the contribution $1 - F_i(t^*)$ to the likelihood function. A spell is left-censored if person i is in poverty at the beginning of the panel, and may have been in this state for a long time. Amemiya (1985) shows that the contribution to the likelihood function then is $[1 - F_i(t)] / \int s f_i(s) ds$.[15] For the sample under scrutiny Table 8.2 collects information on the incidence and duration of poverty spells, and the extent of censoring.

The problem, of course, is how to parametrise the hazard rate $\lambda_i(t)$. A parametrisation, popular in the econometrics of labour turnover, is a Cox proportional hazard rate $\lambda_i(t) = h(t)\exp(\beta' x_i(t))$, where $x_i(t)$ is a vector of time-varying exogenous variables. Following the previous Markov model, $x_i(t)$ includes two different processes: an unemployment process and a household formation process which traces the evolution of the size of the household. Note that the parameter β does not vary with the number of spells. The baseline hazard rate $h(t)$ captures duration dependence of the poverty process.

[15] In some applications, such as duration models of criminal recedivism or fertility, the probability of eventual "failure" is less than one; some censored observations will never "fail". If the survival function is thus defective for some persons, Schmidt and Witte (1989) propose to use a split population model, which parametrises Pr{never fail}$=1 - G(\alpha' z_i) = 1 - 1/(1 + \exp(\alpha' z_i))$, where z_i is a vector of explanatory variables. The likelihood function then needs to be adjusted accordingly.

In the current model, the problem is minor, since this criticism could at most be applied to the old, living on social benefits. However, the density estimates reported in chapter 6 show that poverty is not a predominant old age phenomenon. Moreover, a poor old pensioner could alter her income state by entering the household of her children. The problem, however, is not completely absent given the previous results of the mover-stayer model.

length [years]	numbers of spells	left censored	right censored
1	893	194	145
2	246	62	63
3	113	19	53
4	64	16	27
5	33	6	23
6	30	10	20
7	51	51	51

Table 8.2: Poverty spells: incidence, duration, and censoring

The parameters are estimated by maximising the (partial) likelihood function[16], but in order to simplify the estimation problem left censored spells were deleted. In order to evaluate the integral in (8.5) with discrete data, the exogenous variables were assumed to remain constant during the interval between observations.

The estimation results on the unweighted data are reported in Table 8.3.

Both the nationality and the age variable are not significant. An increased household size increases the poverty hazard. But most important, confirming the evidence of the preceding section, is the employment process. Being unemployed reduces the hazard of leaving poverty.

Furthermore, the plot of the baseline hazard rate $h(t)$ is derived

[16]See Cox (1975) or Lancaster (1990) chapter 9 for a discussion of the partial likelihood function.

variables	ML estimates	standard errors
unemployment status	-.236	.0858
household size	.065	.029
nationality	-.1556	.091
age in 1984	-.0009	.0027

Table 8.3: The continuous-time Cox poverty hazard model

by setting the parameter values β to zero. Inspecting the (not provided) plot of the non-parametric estimate of the baseline hazard rate, it increases at first, but then falls monotonically. Thus medium and long term poverty profiles differ with the latter exhibiting negative duration dependence (see also the non-parametric estimates of the discrete duration model) [17]. Thus, the longer the poverty spell, the less likely is the person to escape from it. However, these findings must be considered tentative in the light of a result due to Heckman and Singer (1984). They have demonstrated that variable selection is a grave problem since "uncontrolled unobservables bias estimated hazard rates towards negative duration dependence". This follows since more mobile persons leave the less mobile persons behind, creating the appearance of stronger negative duration dependence than

[17] $h(t)$ is often assumed to be Weibull, $h(t) = \alpha t^{\alpha-1}$, since a Weibull specification leads to a non-constant hazard rate (but nests within it the exponential distribution which exhibits a lack of memory). Duration dependence is negative (positive) if $\alpha < 1$ ($\alpha > 1$). Fitting a Weibull distribution leads to an estimate of $\hat{\alpha} = 0.833 < 1$, confirming the conjectured negative duration dependence.

actually exists.

Unobservable heterogeneity can be modelled by introducing a mixing distribution, so that the hazard rate is perturbed by an unobservable random variable V. Following Lancaster (1979), let v be iid from a Gamma$(1, \eta)$ distribution with variance η^{-1}, assumed to mimic the unobservables. Thus, the hazard rate for person i becomes

$$\lambda_i(t) = v_i \mu_i(t) \text{ where } \mu_i(t) = \alpha t^{\alpha-1} \exp(\beta' x_i(t)) \qquad (8.6)$$

This specification leads to conditional distributions $F_i(t|v)$ and the unobservable v needs to be integrated out. This yields the unconditional distribution $F^*(t) = E_v(F_i(t|v)) = 1 - [1 + z(t)/\eta]^{-\eta}$ and density $f^*(t) = \mu(t)[1 + z(t)/\eta]^{-(1+\eta)}$ where $z(t) = \int_0^t \mu(s)ds$. The maximum likelihood estimation of the parameter vector (β, η) is carried out using the $EM-$algorithm (see the appendix for a description). However, the resulting estimate of the variance of the mixing Gamma distribution, η^{-1}, is already very high on only the uncensored data, $\eta^{-1} = 14$. This implies that a Gamma mixing model, popular in the literature, is inappropriate in the present context.

It may be argued that the continuous time model is misspecified in that discrete-time data have inappropriately been treated as if they were continuous. Does a discrete-time model have different implications ?

The theory outlined above extends in a natural manner to the discrete-time case. For instance, the hazard rate now has the interpretation $\lambda_i(t) = \Pr\{T_i = t | T_i \geq t; x_i(t)\}$. As pointed out by Allison

(1982) and reiterated by Jenkins (1996), estimation of this model is straightforward. Making the unit of analysis the spell month and thus reorganising the data, the likelihood function for the discrete-time duration model can be rewritten in a form which is standard in the analysis of a binary variable. Two parametrisations of the hazard rate are examined. First, the complementary log-log hazard rate $\lambda_i(t) = 1 - \exp\{-\exp\{h(t) + \beta' x_i(t)\}\}$ is chosen, since it is the counterpart of the underlying continuous time proportional hazard model examined above. But since there is no reason why hazard rates should be proportional, the second parametrisation is the logistic hazard rate $\lambda_i(t) = 1/(1 + \exp\{-h(t) - \beta' x_i(t)\})$.

The results of the estimation are reported in Table 8.4. The selected variables are the same as in the previous models. Duration dependence is captured by the baseline hazard $h(t)$, which is estimated non-parametrically by a sequence of dummies. The results of the two parametrisations are very similar. This should not be too surprising, since it is well known that the logistic model converges to the proportional hazard model as the hazard rate converges to zero. Once again, poverty spells of the long-term poor exhibit negative duration dependence. The hazard of leaving poverty is lower for foreign nationals, and the household formation process is neither important nor very significant. Finding employment is the principal way of escaping from poverty.

What are the determinants of re-entering poverty ? Applying a duration model to this issue is problematic, since sample sizes are

model	c. log-log		logistic	
	ML estimates	SE	ML estimates	robust SE
duration=2 years	.138	.102	.169	.124
duration=3 years	-.163	.159	-.181	.185
duration=4 years	-.549	.232	-.64	.263
duration≥5 years	-.639	.504	-.73	.553
employment status	.407	.0818	.495	.099
household size	.0656	.0283	.0828	.034
nationality	-.161	.089	-.202	.105
disabled	-.044	.092	-.048	.108
education in 1984	.03	.021	.0377	.0268
age in 1984	.0014	.0027	.0018	.0032

Table 8.4: The discrete-time duration models of the hazard of leaving poverty

small: there are 1050 single spells out of poverty which followed a poverty spell of which 70% are right censored. So the subsequent statistical analysis has to be regarded as tentative. However, this data structure suggests that for economically mobile persons (the movers in section 4) poverty is a predominantly transitory and rare event, which once overcome is unlikely to be experienced again.

The hazard rate $\lambda_i(t)$ now captures the probability of person i re-entering poverty at time t. The results of estimating the model with the two hazard parametrisations are reported in Table 8.5. The selected variables are those of the previous models. The estimates show the expected strong negative duration dependence: the longer the spell out of poverty, the less likely is the person to experience poverty again. The surprise, however, is that although the coefficients on all other explanatory variables have signs consistent with the previous results, they are not statistically significant. Moreover, the size of the employment coefficient is very small. This duration model is thus inadequate for analysing the probabilities of re-entering poverty.

8.5.3 German poverty dynamics in context

How do these findings relate to results found by other researchers for other countries such as the US ? The results are, in many ways, similar to those of Bane and Ellwood (1985). Using the PSID for the years 1970 to 1982, they find that most of those who become poor will

model	c. log-log		logistic	
	ML estimates	SE	ML estimates	robust SE
duration=2 years	-.522	.139	-.676	.176
duration=3 years	-.88	.223	-1.1	.263
duration=4 years	-1.11	.279	-1.364	.317
duration≥5 years	-5.12	.326	-5.47	.33
employment status	-.039	.129	-.064	.166
household size	-.043	.045	-.059	.058
nationality	.07	.14	.0999	.185
disabled	.144	.135	.188	.177
education in 1984	-.005	.04	-.0076	.053
age in 1984	-.005	.04	-.004	.005

Table 8.5: The discrete-time duration models of the hazard of re-entering poverty

have only a short stay in poverty, whilst the stock of the poor is predominantly composed of the long-term poor. The hazard of leaving poverty also exhibits negative duration dependence (although these are computed ignoring observable and unobservable population heterogeneity). Using cross-tabulation techniques, they find that earnings changes explain 75% of all poverty spell endings, but this figure is dramatically lower for beginning spells. This result is mirrored in the German case by the importance of the (un)employment process.

The household formation process is found to be of lesser importance than in the US. However, this process is modelled only crudely here as a change in the size of the household, whereas Bane and Ellwood examine separately the various possibilities such as the birth of a child, the wife becoming household head or escaping poverty through marriage or the departure of children from the household. Thus, changes in the size of the household subsume possible events with opposite effects on the poverty status which might explain the small estimated coefficient. Since the current study analyses annual income data, the caveat of Ruggles and Williams (1989) applies, who, using monthly data, find that the typical poverty spell is much shorter than would be anticipated using annual data as 2/3 succeed in escaping poverty before 12 months.[18] In this case, annual data

[18]See also Blank (1989) for a similar econometric approach in the context of single AFDC spells (amongst female household heads) in the US using monthly data. The principal focus of her analysis is duration dependence whose various parametrisations she compares with a non-parametric step-wise specification. In

combines multiple short spells into one long spell, giving the impression of longer duration dependence.

8.6 Conclusion

Intra-distributional mobility is a very important dimension of income dynamics and examining merely the shape dynamics of the income distribution is likely to result in misleading welfare judgments. In the German case, the lack of action at the surface conceals substantial movements beneath it. Indeed according to Friedman's criterion Germany has become a more unequal society because overall mobility has fallen.

Several statistical models based on transition matrices were estimated in order to provide a concise description of the mobility process. The transition probabilities vary with time and the process exhibits a memory which extends beyond one period. The mover-stayer models also suggest the importance of population heterogeneity.

In order to examine the economic determinants of the income process further and to go beyond the descriptive analysis, we have concentrated on one very important income state -poverty- instead of the entire transition matrix. Although different models were estimated - a Markov model with exogenous variables and several dura-

contrast to annual data which might combine multiple short spells, she finds evidence of only weak duration dependence but two distinct groups of beneficiaries.

tion models- the principal findings are similar: unemployment is the principal determinant of poverty; in contrast to the US, the household formation process is only of minor importance, as are age and educational background. Poverty spells of the long-term poor exhibit negative duration dependence: the longer the poverty spell, the less likely is the person to escape poverty.

8.7 Appendix: The EM-Algorithm

This section describes the EM-algorithm used for maximum likelihood estimation in the poverty hazard model with unobservable heterogeneity. For a more detailed description see Cox and Oakes (1984) or Lancaster (1990), upon which the following discussion is based.

The EM-algorithm consists of two principal steps, viz. taking an Expectation, and Maximising the objective function thereafter.

Let the random variable V with realisation v be iid with distribution function $G(.;\eta)$ and associated density $g(.;\eta)$, known up to a parameter vector η. This process generates the unobservable heterogeneity. Let T with realisation t denote the random variable waiting time, parametrised such that its conditional density is $f(t|v;\beta,\eta) = v\mu(t;\beta)\exp(-vz(t;\beta))$ where $z(t;\beta) = \int_0^t \mu(s)ds$. The log-likelihood of the joint distribution of V and T is

$$\begin{aligned}
\log L(\beta;\eta;t;v) &= \sum_{i=1}^{N}[\log f(t_i|v_i;\beta) + \log h(v_i|\eta) \\
&= \sum_{i=1}^{N}[\log v_i + \log \mu(t_i;\beta) - v_i z(t_i;\beta) + \log g(v_i;\eta)
\end{aligned}$$

The algorithm proceeds as follows:

1. From an initial guess $(\beta_n;\eta_n)$ calculate the log-likelihood function of the joint distribution of V and T.

2. Calculate its expected value using the initial guess

$$Q((\beta;\eta)\,(\beta_n;\eta_n)) = E(\log L(\beta;\eta); V|t,(\beta_n;\eta_n))$$

In the context of the present model, the following calculations are typical. Since $g(v)$ and $f(t|v)$ are known, $f(v|t)$ can be calculated. The terms such as $E(V|t)$ and $E(\log V|t)$ are then readily derived.

3. Maximise this with respect to $(\beta;\eta)$. The solutions to the first order conditions define the new iteration values $(\beta_{n+1};\eta_{n+1})$.

4. Continue to iterate until the value of the unconditional log-likelihood converges.

Bibliography

Agell, J. and P. Lundberg (1992). Fair wages, involuntary unemployment and tax policies in the simple general equilibrium model. *Journal of Public Economics 47*(3), 299–320.

Aghion, P. and P. Bolton (1995). A trickle-down theory of growth and development with debt-overhang. *LSE FMG discussion paper No. 170.*

Allison, P. (1982). Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological Methodology 1982.* Jossey-Bass.

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press.

Anderson, T. and L. Goodman (1957). Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 89–109.

Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory 2*, 244–263.

Atkinson, A. (1989). Private and social insurance and the contributory principle. In N. Barr (Ed.), *Current Issues in the Economics of Welfare.* Macmillan.

Atkinson, A. (1990). Institutional features of unemployment insurance and the working of the labour market. *LSE Welfare State Programme discussion paper.*

Atkinson, A. (1993a). On targeting social security: Theory and Western experience with family benefits. *LSE Welfare State Programme discussion paper.*

Atkinson, A. (1993b). What is happening to the distribution of income in the UK ? *LSE Welfare State Programme discussion paper WSP/87.*

Atkinson, A. (1996). Seeking to explain the distribution of income. In J. Hills (Ed.), *New Inequalities.* Cambridge University Press.

Atkinson, A., F. Bourguignon, and C. Morrisson (1992). *Empirical Studies of Earnings Mobility.* Harwood Academic Publishers.

Atkinson, A. and J. Micklewright (1991). Unemployment compensation and labour market transitions: A critical review. *Journal of Economic Literature 29*(4), 1679–1727.

Atkinson, A., L. Rainwater, and T. Smeeding (1994). *Income Distribution in OECD countries: The Evidence from the Luxemburg Income Study.* OECD, Paris.

Bane, M. and D. Ellwood (1985). Slipping into and out of poverty: The dynamics of spells. *Journal of Human Resources.*

Banerjee, A. and A. Newman (1991). Risk-bearing and the theory of income. *Review of Economic Studies 58*(2), 211–235.

Banerjee, A. and A. Newman (1993). Occupational choice and the process of development. *Journal of Political Economy 101*(2), 275–298.

Banks, J. and P. Johnson (1993). Children and household living standards. *Institute for Fiscal Studies.*

Beach, C., K. Chow, J. Formby, and G. Slotsve (1994). Statistical inference for decile means. *Economics Letters 45*, 161–167.

Beach, C. and R. Davidson (1983). Distribution-free statistical inference with Lorenz curves and income shares. *Review of Economic Studies*, 723–735.

Besley, T. (1990). Means-testing versus universal provision in poverty-alleviation programs. *Economica* (57), 119–129.

Besley, T. and S. Coate (1992). Workfare versus welfare incentive arguments for work requirements in poverty-alleviation programs. *American Economic Review* (82), 249–261.

Binmore, K., A. Rubinstein, and A. Wolinsky (1986). The Nash-bargaining solution in economic modelling. *Rand Journal of Economics* (17), 176–188.

Bird, E. (1991). Income variation among West German households. *in Rendtel and Wagner (1991).*

Bishop, Y., S. Fienberg, and P. Holland (1988). *Discrete Multivariate Analysis.* MIT Press.

Blank, R. (1989). Analysing the length of welfare spells. *Journal of Public Economics 39*, 245 – 273.

Boskin, M. and F. Nold (1975). A Markov model of turnover in aid to families with dependent children. *Journal of Human Resources 10*(4), 467–481.

Buchinsky, M. and J. Hunt (1995). Wage mobility in the United States. *NBER working paper 5455.*

Buhmann, B., L. R. N. L. Schmaus, and T. Smeeding (1988). Equivalence scales, well-being, inequality and poverty: Sensitivity estimates across ten countries, using the Luxemburg income study database. *Review of Income and Wealth 34*(2), 115–142.

Burkhauser, R., A. Crews, M. Daly, and S. Jenkins (1995). Where in the world is the middle class ? A cross-national comparison of the shrinking middle class using kernel density estimates. *Syracuse University cross-national studies in aging Discussion Paper 26.*

Burkhauser, R., J. Merz, and T. Smeeding (1994). Relative inequality and poverty in Germany and the United States using alternative equivalence scales. *Discussion paper, Universitaet Lueneburg, Forschungsinstitut Freie Berufe.*

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *Journal of the American Statistical Association 89*(428), 1290–1302.

Champernowne, D. (1973). *The distribution of Incomes between persons.* Cambridge University Press.

Chung, K. (1960). *Markov Chains with stationary transition probabilities.* Springer Verlag.

Coulter, F., F. Cowell, and S. Jenkins (1992). Equivalence scale relativities and the extent of inequality and poverty. *The Economic Journal 102*, 1–16.

Cowell, F. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics 42*, 27–41.

Cowell, F., S. Jenkins, and J. Litchfield (1996). The changing shape of the UK income distribution: Kernel density estimates. In J. Hills (Ed.), *New Inequalities.* Cambridge University Press.

Cowell, F. and M. Victoria-Feser (1996). Robustness properties of inequality measures: The influence function and the principle of transfer. *Econometrica 64*, 77–101.

Cowell, F. and M. Victoria-Feser (1996a). Robustness properties of inequality measures: The influence function and the transfer prnciple. *Econometrica 64*, 77–101.

Cowell, F. and M. Victoria-Feser (1996b). Welfare judgements in the presence of contaminated data. *LSE DARP distributional* *analysis discussion paper No. 13.*

Cox, D. and D. Oakes (1984). *Analysis of Survival Data.* Chapman and Hall.

Dalton, H. (1920). *Some Aspects of the Inequality of Incomes in Modern Communities.* Routledge.

Datenreport 1994. Bundeszentrale fuer Politische Bildung.

Davidson, A. and D. Hinkley (1995). *Bootstrap Methods and their Applications.* Course Notes at Oxford University.

Davidson, R. and J. Duclos (1995). Statistical inferences for the measurement of the incidence of taxes and transfers. *Universite Laval CREFA Discussion Paper No. 9521.*

Deaton, A. (1989). Rice distribution in Thailand. *The Economic Journal supplement*, 1–37.

Department of Social Security (1993). *Households Below Average Incomes.* HMSO.

Diamond, P. (1982). Wage determination and efficiency in search equilibrium. *Review of Economic Studies 49*, 217–227.

Diamond, P. and J. Mirrless (1986). Pay-roll tax financed social insurance with variable retirement. *Scandinavian Journal of Economics 88*(1), 25–50.

Dreze, J. (1990). Famine prevention in India. In J. Dreze and A. Sen (Eds.), *Hunger: Economics and Politics.* Oxford University Press.

Dunford, N. and J. Schwartz (1957). *Linear Operators.* Wiley and Sons.

Easley, D., N. Kiefer, and U. Possen (1985). An equilibrium analysis of optimal unemployment insurance and taxation. *Quaterly Journal of Economics 100*(5), 989–1010.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association 82*, 171 – 200.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap.* Chapman and Hall.

Ferreira, F. (1995). Roads to equality: Wealth distribution dynamics with public-private capital complementarity. *LSE Theoretical Economics Discussion Paper TE/95/286.*

Foster, J., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica 52*(3), 761–766.

Frick, J. and others (1995). Income inequality in East and West Germany before and after unification. *DIW Berlin mimeo.*

Friedman, M. (1962). *Capitalism and Freedom.* Chicago, University of Chicago Press.

Frydman, H. (1984). Maximum likelihood in the mover-stayer model. *Journal of the American Statistical Association 79*(387), 632–638.

Futia, C. (1982). Invariant distributions and the limiting behaviour of Markovian economic models. *Econometrica 50*(2), 377–

408.

Galor, O. and J. Zeira (1993). Income distribution and macroeconomics. *Review of Economic Studies 60*(1), 35–52.

Geweke, J., R. Marshall, and G. Zarkin (1986). Mobility indices in continuous time Markov chains. *Econometrica 54*(6), 1407–1423.

Goodman, L. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association 56*, 841–868.

Gosling, A., S. Machin, and C. Meghir (1996). What has happened to the wages of men since 1966? In J. Hills (Ed.), *New Inequalities*. Cambridge University Press.

Habib, L. (1995). Social insurance in France. *LSE Welfare State Programme Research Note 30*.

Haerdle, W. and O. Linton (1994). Applied non-parametric regression. *Cowles Foundation discussion paper*.

Hall, P., T. DiCiccio, and J. Romano (1989). On smoothing and the bootstrap. *The Annals of Statistics 17*(2), 692–704.

Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel (1986). *Robust Statistics - The Approach based on Influence Functions.* Wiley and Sons.

Heckman, J. and B. Singer (1984). Econometric duration analysis. *Journal of Econometrics*, 32–132.

Hoeffding, W. (1948). A class of statistics with asymptotic normal distribution. *Annals of Mathematical Statistics*, 293–325.

Howes, S. (1993). *Income Distribution.* Ph. D. thesis, London School of Economics.

Huber, P. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, 221–233.

Jahrbuch der Bundesrepublik Deutschland (1994). dtv 5618.

Jenkins, S. (1994). Winner and losers: A portrait of the UK income distribution during the 1980s. *University College of Swansea DP 94-07.*

Jenkins, S. (1995). Did the middle class shrink during the 1980s? UK evidence from kernel density estimates. *Economics Letters*, 407–413.

Jenkins, S. (1996). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics 57*(1), 129–137.

Johnson, S. (1755). *Preface to Dictionary.*

Kendall, M. and A. Stuart (1977). *The Advanced Theory of Statistics, Vol1.* Charles Griffin and Co Ltd.

Klein, T. and G. Zimmermann (1991). Zur oekonomischen Mobilitaet von Individuen und Familien. *in Rendtel and Wagner*

(1991).

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* (4), 939–957.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Econometric Society Monographs 17, Cambridge University Press.

Loury, G. (1981). Intergenerational transfers and the distribution of earnings. *Econometrica 49*(4), 843–867.

Lucas, R. and N. Stokey (1989). *Recursive Methods in Economic Dynamics.* Harvard University Press.

McCall, J. (1970). The economics of information and job search. *Quarterly Journal of Economics 84*(1), 113–126.

McCall, J. (1971). A Markovian model of income dynamics. *Journal of the American Statistical Association 66*(335), 439–447.

Mood, A., F. Graybill, and D. Boes (1974). *Introduction to the Theory of Statistics.* McGraw-Hill.

Piketty, T. (1992). Imperfect capital markets and the persistence of initial wealth inequalities. *LSE Theoretical Economics Discussion paper TE92255*.

Pischke, J. (1994). Continuous training in Germany. *MIT mimeo*.

Pischner, R., U. Rendtel, and G. Wagner (1995). Sampling and weighting of the GSOEP. *DIW mimeo*.

Pissarides, C. (1990). *Equilibrium Unemployment Theory.* Oxford Blackwell.

Poynter, R. and C. Martin (1995). *Rights Guide to Non-Means-Tested Benefits* (18 ed.). Child Poverty Action Group.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery (1992). *Numerical Recipes in C.* Cambridge University Press.

Pudney, S. (1992). Income and wealth inequality and the life-cycle. *Journal of Applied Econometrics 8*, 249–276.

Quah, D. (1994). Convergence empirics across economies with (some) capital mobility. *LSE mimeo.*

Quah, D. (1995). Aggregate and regional disaggregate fluctuations. *LSE Econometrics DP EM/95/290.*

Quah, D. (1996). Empirics for growth and distribution: stratefication, polarisation and convergence clubs. *LSE Department of Economics mimeo.*

Rao, B. P. (1983). *Nonparametric Functional Estimation.* Academic.

Rao, C. (1973). *Linear Statistical Inference and its Application.* John Wiley and Sons.

Rendtel, U. (1995). Documentation of sample size and panel attrition in the GSOEP. *DIW mimeo.*

Rendtel, U., R. Langeheime, and R. Bernsten (1992). The estimation of poverty-dynamics using different household income measures. *Working Papers of the ESF Network on household panel studies. Paper 50, University of Essex.*

Rendtel, U. and J. Schwarze (1991). Die Entwicklung individueller Arbeitseinkommen. *in Rendtel and Wagner (1991).*

Rendtel, U. and G. Wagner (1991). *Lebenslagen im Wandel: Zur Einkommensdynamik in Deutschland seit 1984.* Campus.

Riebschlaeger, M. (1995). A review of weighting methods employed by panel studies included in the PACO project. *CEPS mimeo.*

Ruggles and Williams (1989). Longitudinal measures of poverty: accounting for incomes and assets over time. *Review of Income and Wealth 35*, 225 – 243.

Savin, N. (1984). *Multiple Hypothesis Testing.* Handbook of Econometrics.

Schluter, C. (1995). Social security in Germany. *LSE Welfare State Programme Research Note 29.*

Schluter, C. (1996a). Income distribution and inequality in Germany - Evidence from panel data. *LSE DARP distribution analysis research programme discussion paper No. 16.*

Schluter, C. (1996b). Income mobility in Germany - Evidence from panel data. *LSE DARP distributional analysis research programme discussion paper No. 17.*

Schluter, C. (1997). On the performance of social benefit systems.

Economic Journal 107.

Schmidt, P. and A. Witte (1989). Predicting criminal recidivism using 'split population' survival time models. *Journal of Econometrics 40*(1), 141–159.

Shao, J. and D. Tu (1995). *The Jacknife and the Bootstrap.* Springer Verlag.

Shapiro, C. and J. Stiglitz (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review 74*(3), 433–444.

Shorrocks, A. (1978). The measurement of mobility. *Econometrica 46*, 1013 – 1024.

Silverman, B. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society B 43*, 97–99.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

Silverman, B. and G. Young (1987). The bootstrap: To smooth or not to smooth ? *Biometrika 74*(3), 469–479.

Singer, B. and S. Spilerman (1974). The representation of social processes by Markov models. *American Journal of Sociology 81*(1), 1–54.

Singer, B. and S. Spilerman (1976). Some methodological issues in the analysis of longitudinal surveys. *Annals of Economic and*

Social Measurement (5/4), 447–474.

Stiglitz, J. (1987). The dependence of quality on price. *Journal of Economic Literature 25*, 1–47.

Stiglitz, J. and A. Weiss (1987). Macro-economic equilibrium and credit rationing. *NBER working paper 2164*.

Trede, M. (1995). Statistical inference in mobility measurement: Sex difference in earnings mobility. *University of Cologne, mimeo*.

Veum, J. (1992). Accounting for income mobility in the US. *Social Science Quarterly*, 773–785.

Victoria-Feser, M. (1993). Robust estimation of personal income distribution models. *LSE DARP distributional analysis discussion paper No. 4*.

Yaari, M. (1976). A law of large numbers in the theory of consumer's choice under uncertainty. *Journal of Economic Theory 12*(2), 202–217.

Young, N. (1988). *An introduction to Hilbert space.* Cambridge University Press.