Employment Fluctuation in Rural India : A Statistical Analysis with Special Reference to Maharashtra

Kaushik Bhattacharya

A Dissertation Submitted to the
Indian Statistical Institute
in Partial Fulfilment of the Requirements for
the Award of the Degree of
Doctor of Philosophy

Indian Statistical Institute
Calcutta
1995
Revised Version Submitted : March 1997

Contents

List of Tables

List of Figures

Preface

This dissertation attempts to find out the nature and the extent of employment fluctuation in rural India. The dissertation is partly theoretical and partly empirical in nature. On the methodological side, it builds up a consistent analytical framework. Employment fluctuation is measured in terms of a new mobility measure proposed in this dissertation. The empirical work is based on the data collected by the National Sample Survey Organisation (NSSO), Government of India, in their 38-th round survey operation. I, therefore, thank NSSO for allowing me to work with the data. The views expressed in this dissertation or the errors that remain are, of course, mine and do not reflect that of the Organisation. In this context, I must also thank Prof. Pravin Visaria for his active role in sending the NSSO 38-th round data for rural Maharashtra to me through Prof. Dipankor Coondoo.

The dissertation was prepared under the supervision of Prof. Dipankor Coondoo himself. This, therefore, is a nice opportunity to thank him for his excellent academic guidance at every stage of my research career. But academic guidance is a thing one generally takes for granted from one's supervisor. More than that, I also thank him for the extreme patience with which he tolerated an extremely impatient and nervy character like me for five long years.

I also fondly recall the help provided by Prof. Satya Ranjan Chakravarty and Dr. Amita Majumder. Prof. Chakravarty was always available for discussion and took a very active interest in my research work. Various comments and suggestions from him substantially improved the exposition. On the other hand, Dr. Majumder helped me a lot with my computer work and introduced me to some very useful computer packages. In fact, a part of this dissertation – Chapter 6 to be precise – was done jointly with Dr. Majumder and Prof. Coondoo. I must thank them for their permission to include the work in this dissertation.

Among other academic luminaries of this institute, I gratefully acknowledge the help provided by Professors Nikhilesh Bhattacharya, Mihir Rakhsit, B.V. Rao and T.J. Rao. In spite of their busy schedule, they patiently listened to some of my problems and offered various suggestions. My friends Amar Mukherjee, Pradipta Banerjee and Nirupam Sarkar helped me a lot with my computer work and also in processing this document. Their expertise in their respective subjects reduced my own troubles to a large extent. My friends Murari Mitra and Subhasis Ghosal were also of great help. Discussions with them at various points of time led to several

clarifications in this dissertation.

Finally, I must heartily thank all my friends and contemporaries of this institute. Research is a sequence of hope and frustration and all of us, research scholars, go through these phases from time to time. It is during these phases of frustration that the help of one's friends matters. One need not offer verbal sympathies. A tacit understanding is enough. All my friends at I.S.I. helped me this way indirectly. Their academic devotion, their determination and their way of tackling frustration were infectious. Besides economics and statistics, I have learnt many other things from them and over the years have become a more mature person.

Kaushik Bhattacharya

Chapter 1

Measurement of Unemployment : A Conceptual Framework

1.1 Introduction

Employment is one of the most important socioeconomic variables. Data on employment provide information about the extent of utilisation of the economically active population and are essential for a better understanding of the economic problems of a country. Besides the level of employment, it is also absolutely important to find out the nature and the extent of *fluctuation* of employment for at least three reasons. First, it provides information on *duration* and *spell* of unemployment, i.e., how the total unutilsed labour time of a population within a certain period is distributed across the population (Akerlof and Main, 1980; Bowers, 1980; Hasan and de Broucker, 1982). This information plays an important role in the formulation of the government policy regarding unemployment doles. Second, employment fluctuation is intimately related to the stability of the labour market. Knowledge of the nature and the extent of this fluctuation helps to

(i) predict the time path of unemployment (David and Otsuki, 1968) *and also* (ii) simulate labour force movements (Denton, 1973). Third, employment fluctuation is associated with employment uncertainty. Employment uncertainty of the members of a household means income uncertainty of the household, which is likely to affect the intertemporal consumption decisions of the household, leading to precautionary saving (Leland, 1968). Recently, precautionary saving has been found to explain many of the long standing consumption puzzles and is considered by many economists to be one of the most important contributors to capital accumulation in an economy (Dynarski and Sheffrin, 1987; Blanchard and Mankiw, 1988; Zeldes, 1989; Caballero, 1990, 1991 ; Deaton, 1992).

It is now more or less widely recognised that due to abject poverty in the rural economies of many developing countries like India, very few people can afford to remain unemployed for a long time. They tend to grab whatever work opportunity they come across and after the completion of the contract, become unemployed again (Sundaram and Tendulkar, 1988; Visaria and Minhas, 1991). Since underemployment is a common phenomenon in the rural sector of India, considerable changes in the employment status of individuals take place even *within* a very short reference period (say, a week). Also, a significant portion of the population, especially women, move frequently in and out of labour force – thus enhancing the extent of fluctuation (Bardhan, 1984).

This dissertation attempts to find out the nature and the extent of employment fluctuation in rural India. The dissertation is partly theoretical and partly empirical in nature. With the help of some statistical models, we shall try to relate employment fluctuation to some important individual and household specific covariates. On the methodological side, we propose a new index for measuring employment fluctuation. The empirical analyses of employment fluctuation presented here are based on the household level data on employment and unemployment collected by the National Sample Survey Organisation (NSSO), Government of India, in their 38-th round survey operations, for the rural sector of the state of Maharashtra.

Measurement of the extent and the severity of employment fluctuation is especially pertinent for the rural population because underemployment is a prominent feature of the rural sector. Any such discussion, however, crucially hinges upon the concepts and the definitions of unemployment used. Statistical analyses also depend upon the nature and the type of the collected data. Essentially for this reason, in this introductory chapter we shall discuss briefly the conceptual problems involved in measurement of unemployment and the concepts adopted by the NSSO in their household surveys on employment and unemployment. In the last section of the present chapter, we shall briefly describe the structure of the present dissertation. Needless to mention, the discussion here should help to specify the kind of statistical analyses that need to be undertaken to study the problem of employment fluctuation within the given framework.

The plan of this chapter is as follows : Section 1.2 provides a brief review on the general conceptual framework for measuring unemployment and underemployment. We next discuss in Section 1.3 the conceptual framework adopted by the NSSO for their all-India household employment and unemployment enquiries conducted every five years. Section 1.4 critically reviews this framework of the NSSO. The concepts adopted by NSSO have certain limitations and may not be ideal for the kind of problems addressed in this dissertation. Therefore, these definitions are slightly modified in this dissertation. Section 1.5 describes these changes. Finally in Section 1.6, the types of problems addressed in this dissertation are discussed.

1.2 The General Conceptual Framework

The conceptual problems involved in measuring unemployment have been discussed by many eminent economists.[1] The task of conceptualising unemployment for operational purposes of measurement and policy formulations has two aspects.

First, some summary measures of unemployment have to be defined. The construction of a measure of unemployment involves two logical steps. First, each person in the sample has to be assigned an appropriate labour force category with respect to a reference period. Next, the individual observations should be combined into a single overall index. Following Shorrocks(1993b), we may refer these two steps as the "identification problem" and the "aggregation problem" respectively.

[1]For a brief survey of this literature, see Krishna (1979) and United Nations (1984).

Second, the merits of different sets of quantitative characteristics used to identify various kinds of unemployment need to be critically examined. There are a number of such quantitative characteristics which are relevant for developing some operational definition of unemployment suitable for countries like India, viz., the choice of reference period, classification of the type of activity, alternative concepts of unemployment etc. In what follows, we shall briefly discuss these characteristics.

1.2.1 Choice of Reference Period

Unemployment should be measured with respect to a reference period. This reference period may be a day, a week, a month or a year. The shorter the time period, easier it is to distinguish meaningfully between employment, unemployment and out of labour force status of an individual. If the time reference is lengthened more than a month, it becomes more difficult to distinguish between these three labour market status in which an individual may be during the reference period.

However, if the reference period is made too short, the chance factors may affect an individual's labour market status considerably. In Section 1.3, we shall see that the NSSO tackles this problem by adopting a *moving* reference period throughout the year. This procedural convention greatly reduces the chance factor and also helps to obtain estimates relatively unaffected by seasonality.

1.2.2 Primary Classification by Type of Activity

At any point of time, individuals in an economy are engaged in various type of activities. The classification by type of activity is one of the most useful and primary classifications of the population. A primary classification covering the whole population provides information on different groups with respect to current economic activity.

The details of this classification varies from country to country, each country adopting its own standards. A guideline for broad classification was provided in "Handbook of Population Census Methods" by the United Nations (1958). The primary activity categories adopted by NSSO are described in Appendix B.

1.2.3 The Flow of Time Unemployment Rate (FTUR)

As a summary measure, economists often characterise the extent of unemployment through unemployment rates. The underlying rationale is to focus attention on the extent of unutilised labour.

An activity is defined to be a *gainful activity* if it increases the national income. During the reference period, a person in the economy may be either in the labour force or not, depending upon whether he/she is (i) engaged in a gainful activity, (ii) searching for a gainful activity *or* (iii) simply willing to accept a job, if available. If any one of the above three conditions are satisfied the person is counted within the labour force.

Let the labour supply of a working person or that of a working population be denoted by L^s and the corresponding labour demand is given by L^d. Then the flow of time unemployment rate, u^t, may simply be defined as the ratio of the difference between labour supply and labour demand *and* labour supply, i.e.,

$$u^t = \frac{(L^s - L^d)}{L^s} \tag{1.2.1}$$

Here, L^s and L^d may be expressed in time units and thus may be natural hours, days, weeks or years. They may also be expressed in some efficiency unit[2]. It may be noted that when flows of labour are measured as flows of labour time, they are not only measured in time units, but are also *flows per unit of time*. Thus, we may have L^s and L^d measuring person hours offered or demanded per day respectively.

Clearly, $u^t = 0$ and $u^t = 1$ represent full employment and full unemployment respectively (either for an individual or for the economy as a whole). The intermediate case of $0 < u^t < 1$ represents underemployment, which should be viewed as incomplete employment.

1.2.4 Alternative Criteria for Classifying a Person as Unemployed

FTUR defined in Subsection 1.2.3 above tells us only about the extent of underutilisation of labour time. The classification of a person as unemployed is done on the basis of four major criteria. These are as follows :

[2]An efficiency unit can be defined as the amount of time required to yield either a constant output or a constant income to the worker or the economy.

(i) **time criterion :** the person is gainfully occupied for less than some 'normal' time unit,

(ii) **income criterion :** the person's income per time unit is less than 'desirable',

(iii) **willingness criterion :** the person is willing to do more work and is either actively searching for work or is simply available for work *if it is offered on terms to which he/she is accustomed,*

(iv) **productivity criterion :** the person is removable from his/her present job in the sense that his/her contribution to the output is less than some 'optimal' standard.

Each of these criteria serves its own specific purpose and none can be regarded as the best. The application of these criteria would often yield three or four different measures of unemployment. Sometimes a combination of two or more of these criteria will yield many more estimates (Krishna, 1979). These measures are complementary to each other and are useful for answering different policy questions. When one calculates the number of unemployed persons with respect to all of these criteria, one gets a complete picture of the labour market situation in the economy.

1.2.5 The Stock of Persons Unemployment Rate (SPUR)

Using the definition of FTUR and one or more of the above four criteria, we may measure unemployment not as a flow of labour time per unit of time, but as a stock of persons at a *particular point of time.* The stock status of a person is determined by his/her flow status over the reference period of the census/survey. Given some cut-off levels (norms) of labour supply (L^{s*}) and of the flow rate of unemployment (u^t), a sample person whose actual labour supply L^s exceeds L^{s*} during the reference period is counted within the labour force whereas a person whose u^t exceeds u^{t*} in the reference period is counted as unemployed. If the total number of such unemployed persons is L^u and the total strength of the work force is L^T, the stock of persons unemployment rate, u^p is defined as,

$$u^p = \frac{L^u}{L^T} \tag{1.2.2}$$

1.2.6 Limitations of Unemployment Rate

Although different types of unemployment rates are easy to compute and interpret, they are recently been criticised because they are extremely parsimonious in the use of readily available information (Shorrocks, 1993b). The unemployment rates are based on a one-time "snapshot" of labour force states and consequently contain no memory of individual employment experiences.

To remove these limitations, sometimes the *mean duration of unemployment* is also supplemented along with unemployment rates (Akerlof and Main, 1980). Although the data on mean duration add a longitudinal dimension to the cross-section element captured by the unemployment rate, it may be defined in a variety of ways (Salant, 1977). Also mean duration takes no account of the dispersion in the distribution of spell lengths.

Recently attempts are being made to develop some more general indices of unemployment (Paul, 1992; Shorrocks, 1992, 1993b). These studies often start by proposing some basic properties that should be satisfied by an unemployment index and from these properties try to characterise a class of unemployment measures. These new classes of unemployment indices may be functions of the conventional unemployment rate, the mean unemployment duration and the degree of dispersion in the distribution of unemployment spells (Shorrocks, 1992).

Although these generalised indices are important and they open up a new vista of research, in this dissertation we shall not use them. This is partly because of some limitations in this new literature and partly because our main interest in this dissertation is somewhat different.

Data on duration of unemployment may be collected in various ways. One may observe interrupted or uninterrupted spells, or one may observe a number of individuals for a specified period of time. In the last case, all unemployment spells that terminate within that period are recorded. Here, the individuals may experience multiple spells of unemployment and there is no direct link between the number of spells and the size of the labour force. One limitation of this new literature on unemployment indices is that it cannot effectively deal with multiple spells of unemployment (Shorrocks, 1993b). It is this limitation which makes the new measures less useful in our case. The data which we use is far from ideal for construction of such indices.

The NSSO in India records the employment history of all individuals for the past seven days. On each day an individual may be employed, unemployed or be out of labour force. As employment fluctuation in the rural sector of India is high, many individuals change states even within so brief a reference period.[3] Even if we restrict our attention to those being consistently in the labour force, each of the individuals may have interrupted, uninterrupted or even multiple spells of unemployment.

Also, in our study on unemployment in the rural sector of Maharashtra, we consider separate measures for employment level and employment fluctuation to tackle the static and the dynamic aspects of employment. Thus the employment situation prevailing in the economy is summarised not by one, but by two types of indices. In our study, the traditional unemployment rates are used to summarise the one-time "snapshots" of the labour force states. On the other hand, the dynamic nature of unemployment is summarised by separate measures of employment fluctuation.

1.3 The Conceptual Framework and the Data Collection Procedure of the NSSO

All the Indian five year plans have stressed that a substantial expansion of employment opportunities should be a major goal of planning. Since at the time of independence quantitative estimates of unemployment were unavailable, the first five year plan was based on fragmentary data on employment available from various secondary sources (Bhattacharya, 1981). As the urgency of collection of data on employment for the whole country on a regular basis was felt, the National Sample Survey (NSS) was set up in 1950 as a part of the Indian Statistical Institute for the collection of various kinds of socioeconomic data for the country as a whole through field surveys for national accounting, planning and other policy purposes. The task of collection of data on employment and unemployment was also entrusted to the NSS. The first comprehensive household survey on employment and unemployment in India was conducted in 1955. Since then the NSS have been conducting quinquennial surveys on employment and unemployment

[3]See Chapter 2 of this dissertation.

in India.[4]

This section provides a brief description of the concepts adopted in the 38-th round survey operation of the NSSO. A brief discussion on the sampling design, survey coverage, estimation procedure and other salient features of the 38-th round survey operation of the NSSO is provided in Appendix A. Details of all these may be found in Report No. 341/4, NSSO, Government of India (1988).

1.3.1 The Conceptual Framework of the NSSO

Over the years, the conceptual framework for the survey of employment and unemployment adopted by the NSSO has undergone many changes (Bhattacharya, 1981). During the 60's, it was felt that the simple measures like the existing stock of unemployed persons or percentage of unemployed persons would not reveal the true situation in the labour market in India as it did in the advanced countries. Estimates were also required for homogeneous segments of the labour force defined in terms of such important characteristics as region, sex, age, education, occupation, class of the worker, rural-urban residence of the worker etc. (Mahalanobis, 1968). Also, it was felt that the sample households should be distributed over the agricultural seasons in a manner so that valid estimates of unemployment rates for each of the season can be obtained separately. The "Report of the Committee of Experts on Unemployment Estimates" (Dantwala Committee) set up by the Planning Commission for an improvement in the data collection scheme was a definite watershed in this process of evolution. Since, we shall work with the household level data collected by the NSSO in their 38-th round survey operations (January – December, 1983), we shall restrict our discussion to the conceptual framework adopted in the 38-th round survey operation only.

Among the various criteria described in Subsection 1.2.4 above, the NSSO adopts a mixture of time criterion and willingness criterion. Since the early 1970's, the NSSO has begun to measure unemployment according to three alternative definitions or concepts. Each definition or concept serves its own specific purpose and identifies different features of the employment

[4]In 1972, the NSS was taken over by the Government of India. It was placed under the Department of Statistics of the Ministry of Planning and was renamed as the National Sample Survey Organisation (NSSO).

situation prevailing in the country. They are, (i) the usual status approach, (ii) the weekly status approach and (iii) the daily status approach. The first of these, the usual status approach, focuses on the chronically unemployed while the second one is concerned with the state of employment of individuals during the reference week of the survey. In addition, the measurement of unemployment in terms of person-days, which takes account of unemployment within the reference week has been an Indian contribution to the subject, now recommended internationally as well (United Nations, 1958; Visaria and Minhas, 1991). It may be noticed that this third concept, in particular, becomes very helpful in measuring underemployment.

Persons belonging to the three employment states are further classified by attributes like age, sex, type of general and technical education, industry, occupation etc. The classification of individuals for these three measures is done according to the activities they pursue for the greater part of a year, a week or one half of a day respectively. The different activity categories used by the NSSO are presented in Appendix B while the corresponding definitions are given in Appendix C.

1.3.2 Classification according to Usual Status

In the 38-th round survey of the NSSO, the status of activity in which an individual spent relatively longer time of the preceding 365 days prior to the date of the survey, was considered to be his/her usual activity status. Accordingly, a person was classified as 'working' or 'employed' if he was engaged for the major part of the preceding year in activity categories coded as 01-51, 'unemployed' for activities coded as 81-82 and 'out of labour force' for activities coded as 91-98 (see Appendix B). Within the two broad activity categories, 'working' and 'out of labour force', the detailed activity category was determined on the basis of *time criterion*. A person categorised as 'worker' on the basis of his/her principal status was referred to as a *principal status worker*. Those of the 'non-workers' who pursued in a subsidiary capacity some gainful activity along with their principal non-gainful activity, were considered to be usually working in a subsidiary capacity and were referred to as *subsidiary status workers*. These two groups, viz., principal status workers and subsidiary status workers together constituted *all workers* according to the usual status classification.

1.3.3 Classification according to Current Weekly Status

For classification of persons according to current weekly status, persons were assigned a unique activity status. For persons pursuing different type of activities approximately for the same time within the reference week, the status of their activities was decided by adopting a priority-cum-major time rule. Under the priority rule, the status of 'working' got priority over the status of 'not working but seeking or being available for work', which in turn got priority over the status of 'not working and not seeking or not being available for work'. For a person classified under the first or the last category, his/her detailed activity status was determined by major-time-spent criterion. Further, if a person had worked for at least one hour on any day of the reference week, he/she was considered as 'working'. A person who had not worked during the week but was available for work even for one hour on any day of the reference week was considered as 'seeking and/or was available for work'. Others constituted the category 'not available for work'. Aggregate of persons under the different activity status provided the distribution of persons by activity on an average in a week of the survey period of one year.

1.3.4 Classification according to Current Daily Status

In adopting the current daily status approach attempt was made to account for the two major activities pursued by a person on each day of the reference week, allocating 'half day' to each activity. The basic unit of classification was thus *half-a-day*. In assigning the activity status on a day, a person was considered 'working' for the entire day if he/she had worked 4 hours or more on the day, and assigned the one or two (as the case might be) work activities (01-71) to which he/she devoted the working time. But if the work was done for 1 hour or more but less than 4 hours, he/she was considered as 'working' (employed) for half day and 'seeking or available for work' (unemployed) or 'not available for work' (out of labour force) for the other half of the day depending upon whether he/she was seeking or was available for employment on the day. On the other hand, if a person was not engaged in any gainful work even for one hour of the day but was seeking or was available for work for four hours or more, he/she was considered 'unemployed' for the entire day. But if the reported availability for work was for less than 4 hours only, he/she was considered 'unemployed' for half day and 'out of the labour force' for the other half of the day. A person who was neither having any work to do nor was available for work even for one half of the day, was considered as 'out of labour force' for that day and assigned the one or two of the non-gainful activity status which he/she had during the day. The aggregate of person-days so classified under the various activity categories for all the seven days of the week divided by seven gives the distribution of persons (strictly speaking, person-days) by activity category on an average day of the one year survey period.

1.3.5 Probing Questions

In order to understand the complex nature of work patterns, in the 38-th round survey two sets of probing questions were asked. One set was administered on all persons aged more than five years except those who were either very old or disabled. The second set concentrated on those who were usually engaged in domestic duties. This was aimed at bringing out the extent of participation of women engaged in household chores to some specialised activities. The questions asked were supposed to reveal the extent of discouraged drop out (Bardhan,1984).

1.4 A Critical Review

The NSSO data relating to unemployment are often questioned on the ground that they indicate a much smaller prevalence of unemployment than poverty. The widespread presumption of a close link between poverty and unemployment overlooks the fact that in the absence of unemployment insurance, very few people in the rural areas can afford to remain idle for a long time. Therefore, the usual status SPUR should not be used as a proxy for the level of living of the households.

The weekly status measurement of unemployment, on the other hand, minimises the probability of a person being classified as unemployed by combining a very liberal definition of employment (requiring only one hour of 'gainful activity' on any day of the reference week). An alternative measure suggested by Visaria(1981) has yielded more pessimistic estimates and has exposed the restrictive nature of this measure.

The current daily status measure, however, is expected to capture the underutilisation of labour time well. With half-a-day as the basic unit, FTUR obtained from this approach, when averaged over all seasons, maybe considered to be a fairly reliable estimate of the average extent of underutilisation of labour time over the whole year. It exhibits more variability across regions, seasons, sex and other important covariates of employment. Even then, it has been suspected that due to the problem of evaluating the response of the self-employed of being 'at work', the person-day unemployment rate (PDUR) reported for them may be an underestimate of the underutilisation of labour time for this group (Sundaram and Tendulkar, 1988). Also, this measure cannot really reflect the level of living unless it is correlated with the wage rates of various occupational groups. Such a task will be a difficult one because the shadow wage rate for the self-employed, who constitute a large part of the total population, is to be determined accurately for this purpose.

Over the years, the data collected by the NSSO have provided more or less reliable estimates and in most cases had been seen to be compatible with a priori expectations. Moreover, with time, the concepts and the data collection procedures have also improved. The 38-th round survey data, a part of which we have used in this dissertation, have been collected after those modifications suggested by various experts were implemented. In this context, it may be mentioned that the results we have obtained do not indicate any abnormal deviations from our a priori perceptions.

1.5 The Concepts Adopted in This Dissertation

In Section 1.2, we discussed that the shorter the time period – say, a day or a week – the easier it is to distinguish meaningfully between employment and unemployment. If the time reference is lengthened to a month or a year, it becomes much more difficult to distinguish between the two states.

Throughout this dissertation we shall use the daily status measurements because only from this measurement the day-to-day fluctuation of employment of individuals can be constructed. However, we have modified the NSSO definition of employment slightly in order to make it compatible with our theoretical framework of analysis. The modifications that have been made is explained below.

From the activity category listed in Appendix B, for any day during the reference week a person may be classified into any of the six following categories :

(i) Fully employed

(ii) Fully unemployed

(iii) Fully out of labour force

(iv) Half employed – half unemployed

(v) Half employed – half out of labour force

(vi) Half unemployed – half out of labour force

This means that on any day of the reference week a person may belong to any of these six *states* and movements are possible across these six states. A part of the research reported in this dissertation has a theoretical framework based on a simple Markov chain model defined on alternative employment states. As such an analysis becomes very complicated if the number of states considered is large, we have combined the above six states suitably to have fewer states.

To be precise, we have defined an individual as employed (E) if he spends four hours or more on a given day in gainful activities. If, however, the individual works less than four hours on a day but searches for work (or, is simply available for work) for at least one hour, he or she is classified as unemployed (U). On the other hand, if the individual works for less than four hours on a day and does not search (or, is not available) for work for at least one more hour, he or she is classified to be out of labour force (O).

Note that this definition differs slightly from the NSSO definition, which considers half-a-day as the basic unit. We have simplified this classification by noting that very few persons actually change their activity status within a day and thus, reduced the number of employment states by suitably regrouping them. In Chapter 2, we shall see that the above simplification changes our measures of employment and unemployment rates from that of the NSSO estimates only very slightly.

All discussions on employment and unemployment should clearly identify the population by explicitly specifying a minimum age limit. As labour market conditions vary from country to country, this minimum age limit should be set in accordance with the condition prevailing in each country. Since child labour is very widely used in the rural sector of India (Paul, 1988), in the present case this minimum age limit should not be very high. However, since babies are always out of labour force, we have taken this minimum age limit to be five years. This minimum age limit tallies with the one generally adopted by the NSSO in their published reports.[5] We have, however, fixed no upper age limit. For the rural sector of India, it is difficult to fix this upper age limit because farmers and self-employed persons may continue to work until they are very old.

Thus the population that we have considered in the empirical study here consists of all persons in rural Maharashtra with age greater than or equal to five years. In some special cases, however, we also study the employment behaviour separately for persons having age greater than or equal to fifteen years. The main purpose of such an analysis is to have an easy cross-validation of our results with those published in the NSSO reports.[6]

Though unemployment based on daily status measure provides a more realistic estimate of the unutilised labour-time, it has its limitations – particularly if the reference period of data collection is short. The first and the foremost consideration is that the chance factors can affect the labour market activities of people during the very short reference period. For example, one day may be a holiday or during one week the weather may be bad and therefore the labour market activities of the people get curtailed drastically. Although the NSSO adopts a moving reference period for data collection, such problems cannot be totally ruled out. Therefore, some of the inferences drawn from our studies may be weak and should be interpreted as hypotheses to be verified from other sources, if possible.

[5]See Report No. 341/4, 'Report on the Third Quinquennial Survey on Employment and Unemployment, Maharashtra', NSSO, 1988.

[6]Op cit.

1.6 The Plan of This Dissertation

The plan of this dissertation is as follows : Chapter 2 presents a brief empirical examination of the extent and the nature of underemployment and employment fluctuation in rural Maharashtra. In this study, we have measured employment fluctuation by the number of transitions (jumps) made by individuals within a given reference period. We examine the variations in underemployment and employment fluctuation separately across some covariates.

Chapter 3 presents some general statistical models which try to explain the variation in the observed number of jumps made by individuals. The models developed are primarily based on Poisson and Negative Binomial distributions. We relate the number of jumps to some important individual and household specific covariates like region, subround, age, sex, caste, religion, education etc. in an attempt to explain the observed variations in the number of jumps for individuals.

Although the number of jumps effectively summarises the extent of employment fluctuation, some amount of information is lost because of this summarisation. Alternatively, the problem of employment fluctuation can be formulated as a problem of transition from one state of employment to another during the reference period. Problems of this type arise quite frequently in different fields of study (e.g., in migration, in job mobility etc.). Chapter 4 tries to analyse the problem of employment fluctuation with the help of a simple Markov chain model which examines the day-to-day transitions across employment states. The transition matrix of this Markov chain is supposed to reveal the extent and the nature of employment fluctuation.

In economics of transition, a problem which occurs frequently and severely limits the scope of analysis is that of aggregation. Often the transition data available are aggregated over individuals, (e.g., one may know only the aggregate number of persons employed, unemployed and out of labour force in each period). In such a case, a researcher is forced to assume some multivariate time series models – the most common form of time series model used being the vector autoregressive moving average (VARMA) model.

Chapter 4 examines the aggregation implication of time series models under the assumption that the day-to-day employments of individuals follow stationary and independent Markov chains. Here we show that the conditional distribution of the cell aggregates follows a nonstationary vector autoregressive (VAR) process, with both its mean vector and transition matrix being functions of the transition probabilities.

Chapter 5 logically connects Chapters 2, 3 and 4. In Chapters 2 and 3, employment fluctuation is measured by the total number of state changes observed for individuals in the reference period. The number of jumps from one employment state to another basically reflects *the extent of movement* in the corresponding stochastic process. Chapter 5, on the other hand, starts from the same individual-specific assumption as in Chapter 4, viz., employment of individuals follow stationary and independent Markov chains.

Chapter 5 presents a measure for employment fluctuation based on the notion of *predictability* of the time path of the aggregate employment vector. The measure we propose is derived from the period-to-period variation of the cell aggregates of the individuals. The measure has some nice probabilistic and information theoretic interpretations. It can be interpreted as a weighted average of the entropies of the probability distributions corresponding to the rows of the transition matrix. Also, it is proportional to the probability that two individuals A and B will be in different states in the next period given that they are in the same state in the current period.

Although the transition matrix of a Markov chain reveals the nature and the extent of transitions, for comparison of fluctuations of two different transition matrices we need to have a *single* measure to be constructed from the transition matrices. These measures are called *mobility indices.* Mobility measures play an important role in summarising the extent of movement and the extent of predictability from a transition matrix. Over the years, many such measures have been proposed and used in economics, sociology and other disciplines.

Chapter 5 presents a brief discussion on mobility indices and shows that the proposed measure of employment fluctuation can also be interpreted as a measure of mobility over an important class of transition matrices that are frequently encountered in reality. The measure is then applied to the NSSO 38-th round data for rural Maharashtra to examine the nature of employment predictability of different subgroups of the population under consideration.

Chapter 6 considers another alternative approach to the study of employment fluctuation. Since underemployment is a common phenomenon in the Indian rural labour market and also, employment fluctuates considerably, households are exposed to the risk of unemployment. The existing literature ignores this problem of variation of employment risk across various types of households. In almost all these studies some aggregate measures of employment is regressed on its determinants. Apparently, adequate attention has not been given to the effect of individual and household specific variables on employment risk.

In Chapter 6, we show that under certain simple individual-specific behavioural assumptions, ordinary least squares (OLS) will not be a valid model. We consider the implications of two simple assumptions : (i) the day-to-day employment of individuals are independent *and* (ii) the day-to-day employment of individuals form independent Markov chains. In both cases we show that under some intuitively appealing specifications, the assumption of constant variance of employment rate across individuals is no longer valid. In fact we show that the distribution of aggregate employment rate of individuals will belong to the class of models proposed by Just and Pope(1978, 1979) which had already been used widely in the context of risk related to agricultural production. Empirically, we test the presence of heteroscedasticity across various covariate groups. The statistical tests categorically reveal that the problem of difference in employment risk cannot be neglected.

Finally, Chapter 7 summarises the main findings and provides some concluding comments and observations.

There are three appendices to this dissertation. Appendix A describes the survey design and the estimation procedure of the NSSO. It also adds some critical comments. Appendix B presents the NSSO definitions of various activity codes and finally Appendix C provides data definitions.

Chapter 2

Underemployment and Employment Fluctuation in Maharashtra : A Descriptive Analysis of the Data

2.1 Introduction

This chapter carries out a preliminary data analysis regarding the extent and the nature of underemployment and fluctuation of employment in rural Maharashtra. It correlates employment fluctuation with some important covariates like geographical area, agricultural season and various socioeconomic groups of the population. This chapter, however, only serves as a building block for the problem of employment fluctuation which is analysed with the help of more sophisticated statistical models in subsequent chapters. Therefore, in this chapter we examine the relationship of each covariate group with underemployment and employment fluctuation *separately* ignoring any interrelationship among the covariates.

The plan of this chapter is as follows : Section 2.2 presents a very brief survey of the main findings on the extent of underemployment in rural Maharashtra. Section 2.3 carries out a descriptive data analysis with the help of the NSSO 38-th round data on employment and unemployment for the State. We calculate unemployment rate across several important covariates and compare our results with previous such findings. In Section 2.4, the extent and the nature of employment fluctuation is examined. In this chapter, we measure employment fluctuation by the number of state changes ('jumps') made by the individuals over the reference period. We evaluate the distribution of jumps for the aggregate data as well as for groups of households classified by some important covariates and use these to identify the groups of the population which are more affected by employment fluctuation. Finally, Section 2.5 summarises the main findings and makes some concluding observations.

2.2 A Brief Review of Underemployment in Rural Maharashtra

Maharashtra is the third largest state in India, both in terms of area and in terms of population. Its geographical area is 307,762 sq. kms. The total population of the State as per 1991 census was 7.88 crores. However, the estimated population of the state, as on July 1, 1983, was 6.59 crores.[1] Out of this total population, approximately 65 per cent were found to reside in rural areas.[2]

The economy of Maharashtra presents a picture of extreme contrasts. The organised industrial sector in this state is concentrated mostly in the three districts of Bombay, Thane and Pune. Outside this industrial belt traditionalism, tribalism and low yields and low productivity continue to prevail. In agriculture too, excluding the pockets of sugarcane areas in the Western part of Maharashtra, the rest of the state is dominated by rainfed traditional agriculture.

In addition to the vulnerability in respect of rainfall, very low percentage of gross cropped area is irrigated. Since most of this cropped area fall in dry regions, the ultimate irrigation potential of the state is not much. Also, there exists significant inter-regional variations in the extent of irrigation. In Western Maharashtra and in the Eastern regions, the percentage of area ir-

[1] This estimate was obtained by interpolating the corresponding population estimates as on March 1, 1981 and March 1, 1986. See page A-320 of Report No. 341/4, 'Report on the Third Quinquennial Survey on Employment and Unemployment, Maharashtra', NSSO, 1988.

[2] Op cit.

rigated varies from 16 to 23 whereas in Konkan, Vidarbha and Marathwada[3] irrigation is less than 10% (Dev, 1992).

In Maharashtra, like most other states in India, more than 80% of the rural workers are engaged in agriculture. In view of the limited facilities, the opportunities for employment throughout the year in the rural areas are limited. Subsistence farming in dry areas and insecurity of gainful employment in the entire rain-shadow areas which are chronically affected by drought lead to considerable unemployment and underemployment in the rural parts of the State. From May 1972 onwards, the Government of Maharashtra introduced an Employment Guarantee Scheme (EGS) throughout rural Maharashtra for mitigating the difficulties that the rural labour force face in finding employment. The EGS provides an unlimited guarantee of employment to all adults above 18 years of age who are willing to do unskilled manual work on a piece rate basis. The scheme aims at sustaining household welfare in the short run and also at building up directly productive community assets.

It is generally believed that the EGS is one of the most successful schemes ever pursued in India in reducing poverty and unemployment (Drèze and Sen, 1990). There is, however, some heterogeneity *within* the state so far as the success of the EGS is concerned (Ezekiel and Stuyt, 1990). Estimates obtained from various sources regarding the extent of EGS employment also vary widely (Dev, 1992).

One of the most important aspects of the EGS is that it has made the rural labour market more stable. It has helped to prevent acute distress and avoid such costly forms of adjustments as selling of assets. In other words, even if the increase in employment and income is not large, the effect of the existence of an employment or an income insurance has been very significant. The study of Walker and Ryan (1990) reveals that the EGS villages have about 50% less variable income stream than the non-EGS villages.

We now present some previous results on underemployment and employment fluctuation. One of the major problems in this context is the lack of a reliable parallel data base which can be used for an external validation

[3]Historically, Konkan, Western Maharashtra, Vidarbha and Marathwada are the four main subdivisions of Maharashtra. The NSSO, however, divides the State into six agroclimatic regions. See Appendix C for the definition of the NSSO regions of Maharashtra and how they are related with this historical subdivision.

of the NSSO estimates. Unemployment by usual status for rural Maharashtra estimated from the 32-nd round NSSO data was 1.61% while the estimated person-day unemployment rate (PDUR) for the same was 7.35% (Paul, 1988). The unemployment rates were found to vary widely across the population. Unemployment and underemployment among females were higher than that of males. The PDUR for males and females were 5.85% and 9.31% respectively. Among occupational categories, the condition of the agricultural labourers were very critical. The PDUR for them was 13.07% and agricultural labour alone contributed about 70% to the total unemployed person-days in the rural areas of the State (Paul, 1988).

2.3 Underemployment in Rural Maharashtra : A Preliminary Data Analysis

In this section, we examine the nature and the extent of underemployment as revealed by the NSSO 38-th round survey data on employment.[4] Altogether the NSSO 38-th round survey on employment and unemployment covered about 568 villages of Maharashtra. The number of sample households surveyed was 5388, with the number of sample persons being 27756. Out of these 5388 households, details of daily status measurements were unavailable for 280 households. Excluding those having age less than five completed years, our sample consists of 23474 persons.

Using the NSSO 38-th round survey data on employment and unemployment, in this section we shall calculate PDUR for the aggregate as well as separately for groups of population classified by some important covariates like region, occupational category, caste and religious group of the households and age, sex and education of the individuals. Thus our study will be similar in nature to that of Paul (1988).

First, we examine the extent of fluctuation in aggregate employment. Over the period 1977-78 to 1983, the usual status unemployment has declined considerably in the rural sector (Paul, 1988). We shall examine whether this is also true for indicators of underemployment based on daily status measures.

[4]For a discussion on the NSSO 43-rd round survey data on employment, unemployment and underemployment, see Ray and Jacob(1990) and Dev(1992).

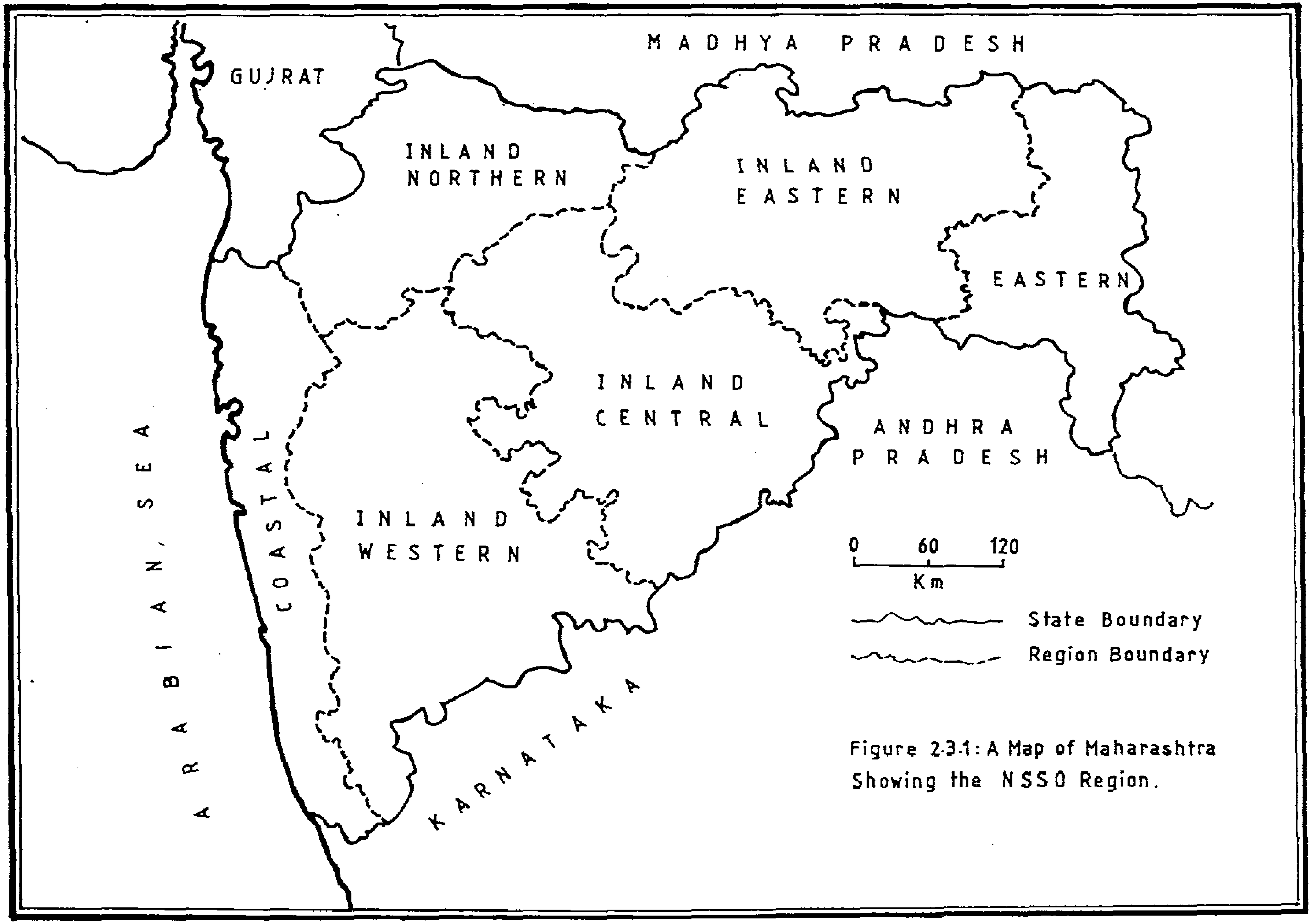

Figure 2.3.1: A Map of Maharashtra Showing the NSSO Region.

2.3.1 Underemployment in the Aggregate

In this subsection, we present the aggregate PDUR separately for the seven days of the reference week. There is, however, a methodological problem involved. The NSSO adopts moving reference period even within a sub-round, i.e., two households which were visited in the same agroclimatic season might have been visited on different days. Since the calendar dates are not available, the aggregates should not be interpreted as pertaining to a specific calendar date. And yet such an exercise is immensely useful because it checks one important aspect of the validation exercise of the NSSO data which is not very often repeated.

The results on day-to-day estimates are presented in Table 2.3.1. In all the tables reported in this chapter, Emp, Unemp and Outlab mean percentages of the total population employed, unemployed and out of labour force respectively. Unemprate denotes the person-day unemployment rate which is defined as :

$$\text{Unemprate} = \frac{\text{Unemp}}{\text{Emp} + \text{Unemp}} \tag{2.3.1}$$

Note that the above unemployment rate is a flow of time unemployment rate defined in Chapter 1.

The NSSO survey design is not self-weighting. Each sample household represents a different number of households in the population. Since these numbers, which are called *multipliers*, are different for different households, all estimates of aggregates or averages in this dissertation are calculated by using the *multipliers* as weights. The details of method of estimation is described in Appendix A.

Table 2.3.1 shows that the average of the unemployment rates across days, which gives us PDUR for the reference week, is 0.0675. Note that the definition of employment adopted by us will yield slightly higher estimates of unemployment rate than that according to the NSSO definition.[5] A comparison of this with the corresponding estimate of 0.0735 of Paul (1988) for the year 1977-78 mentioned in Section 2.2 above suggests that over the period of 1977-78 to 1983, underemployment in the rural sector of Maharashtra has

[5]The estimated PDUR by the NSSO definition is 6.62% (See Table 30, A-124 in Report No. 341/4, 'Report on the Third Quinquennial Survey on Employment and Unemployment, Maharashtra', NSSO, 1988.), which is very close to the estimate according to our definition.

Table 2.3.1: Daywise Estimated Proportion of Persons by the States of Employment and the Unemployment Rate for Seven Successive Days : Rural Maharashtra, NSSO 38-th Round Combined Sample

States of Employment	Estimated proportion of persons for							Average for seven days
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Emp	0.4572	0.4579	0.4588	0.4615	0.4639	0.4690	0.4668	0.4622
Unemp	0.0360	0.0354	0.0351	0.0345	0.0328	0.0300	0.0304	0.0335
Outlab	0.5067	0.5067	0.5060	0.5040	0.5032	0.5010	0.5028	0.5043
Unemprate	0.0730	0.0718	0.0711	0.0695	0.0660	0.0601	0.0611	0.0675

somewhat decreased.

We now examine the nature and the extent of underemployment across some important covariates. We restrict our attention to seven such covariates : (i) region, (ii) agricultural season (subround), (iii) household's main occupational type, (iv) social and religious group, and finally (v) age, (vi) sex and (vii) education of the individuals. Each of these covariates can be broken into several categories. We examine whether PDUR changes substantially across these categories. For convenience, the number of sample persons for each covariate category (Per) is also presented in the tables along with unemployment rate.

2.3.2 Effect of Region

Although Maharashtra is considered to be one of the more developed states in India in terms of several conventional indicators of development, these aggregate indices hide enormous regional differences that exist within the State. It is quite well known that the neighbourhood of Bombay and the Western part of Maharashtra are highly developed. The rest of the State is grossly underdeveloped relative to these parts (Dev, 1992). The NSSO divides the state into six agroclimatic regions, viz., (i) Coastal, (ii) Inland Western (Western), (iii) Inland Northern (Northern), (iv) Inland Central (Central), (v) Inland Eastern (Ineast) and (vi) Eastern.[6] These regions are formed with the hope that separate regionwise estimates of unemployment will provide a better and a clearer picture of the labour market situation in

[6]See Appendix C for the definition of these regions.

Table 2.3.2: Regionwise Estimated Proportion of Persons by the States of Employment and the Unemployment Rate : Rural Maharashtra, NSSO 38-th Round, Combined Sample

States of Employment	Estimated proportion of persons for						All Regions
	Coastal	Western	Northern	Central	Ineast	Eastern	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Emp	0.4300	0.4590	0.4795	0.5016	0.4314	0.4631	0.4622
Unemp	0.0264	0.0236	0.0508	0.0310	0.0378	0.0462	0.0335
Outlab	0.5436	0.5174	0.4697	0.4674	0.5308	0.4907	0.5043
Unemprate	0.0578	0.0489	0.0958	0.0582	0.0806	0.0907	0.0675
Per	2495	6793	3953	3976	4266	1991	23474

Maharashtra than any aggregate measure comprising of the whole State. In Figure 2.3.1, we show the various NSSO regions on a map of Maharashtra and in Table 2.3.2 we present the estimated proportion of persons in different states of employment and the unemployment rate separately for these regions.

Table 2.3.2 shows wide heterogeneity in the rate of unemployment across regions. It clearly identifies the Coastal, Inland Western and the Inland Central parts of Maharashtra as the regions having lower unemployment rates compared to the other regions of the State. The Inland Northern region comprising of the districts of Nasik, Dhule and Jalgaon has the highest unemployment rate (9.58%). The Eastern and the Inland Eastern parts of Maharashtra, which are economically backward, come close to this region.

2.3.3 Effect of Seasonality

The pattern of employment in the rural sector is likely to show marked seasonality due to seasonal variation in the level of agricultural activities. During the 'peak' season, employment opportunity increases considerably compared to the 'lean' season. On the other hand, in the peak season employment fluctuation generally increases because a significant number of women move more frequently in and out of labour force (Bardhan, 1984).

In Table 2.3.3 we present the various estimated proportions relating to employment status by subrounds. To be more specific, in this table the

Table 2.3.3: Subroundwise Estimated Proportion of Persons by the States of Employment and the Unemployment Rate : Rural Maharashtra, NSSO 38-th Round, Combined Sample

States of Employment	Estimated proportion of persons for				All Subrounds
	Subround 1 (Jan-Mar)	Subround 2 (Apr-Jun)	Subround 3 (Jul-Sep)	Subround 4 (Oct-Dec)	
(1)	(2)	(3)	(4)	(5)	(6)
Emp	0.4618	0.4254	0.4818	0.4834	0.4622
Unemp	0.0293	0.0404	0.0371	0.0278	0.0335
Outlab	0.5089	0.5342	0.4811	0.4888	0.5043
Unemprate	0.0597	0.0868	0.0715	0.0545	0.0675
Per	6162	6203	5032	6077	23474

results are presented for the four subrounds of the NSSO enquiry, covering the months January to March (Subround 1), April to June (Subround 2), July to September (Subround 3) and October to December (Subround 4). It may be noted that the lean and the peak season roughly correspond to Subrounds 1–2 and Subrounds 3–4 respectively.

Table 2.3.3 suggests that in spite of the successful implementation of EGS, the unemployment rates for the four subrounds in Maharashtra are not very close, the range of unemployment rate being 0.0323. We also observe that the unemployment rate throughout the periods April to September remains higher than the annual average. As the period April to June falls under lean season, the percentage of time spent outside labour force is the highest. On the other hand, as the period July to September is considered as the beginning of the peak season, the percentage of time spent outside labour force is the lowest.

2.3.4 Effect of Household Type

Following the NSSO definition, the rural households may be classified into five major occupational types according to the major source of income of the respective households for the past 365 days prior to the date of the survey. These are (i) self-employed in agriculture (Selfagr), (ii) self-employed in non-agricultural activities (Selfnagr), (iii) agricultural labourers (Agrlab),

Table 2.3.4: Occupationwise Estimated Proportion of Persons by the States of Employment and the Unemployment Rate : Rural Maharashtra, NSSO 38-th Round, Combined Sample

States of Employment	Estimated proportion of persons for					All Households
	Selfagr	Selfnagr	Agrlab	Othlab	Othocc	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Emp	0.4946	0.3999	0.4646	0.4046	0.3765	0.4622
Unemp	0.0096	0.0200	0.0627	0.0570	0.0206	0.0335
Outlab	0.4958	0.5801	0.4727	0.5384	0.6028	0.5043
Unemprate	0.0192	0.0476	0.1189	0.1234	0.0519	0.0675
Per	10106	1926	8475	1466	1501	23474

(iv) other labourers (Othlab) and (v) others (Othocc).

A priori, one would expect household type to be a major concomitant of the employment status of the members of a household. This is because in the rural sector, employment often depends upon access to an adequate asset base. The self-employed households are, therefore, more protected from the uncertainties in the labour market (Sundaram and Tendulkar, 1988; Paul, 1988; Visaria and Minhas, 1991). Estimates of various employment related proportions by household type are presented in Table 2.3.4.

The results in Table 2.3.4 are fully consistent with the findings of earlier studies. It highlights the wide difference in unemployment across various household types. While the unemployment rate of the households which are self-employed in agriculture is as low as 0.0192, that of the agricultural labourer and other labourer households are 0.1189 and 0.1234 respectively. There is also a wide difference in the percentage of time spent outside labour force. For example, while for the self-employed households in non-agricultural occupations and for others these figures are more than 58% and 60% respectively, it is only 47.27% for persons belonging to the agricultural labourer households. This implies that the agricultural labourers cannot afford to remain idle for a long time.

Table 2.3.5: Social Groupwise Estimated Proportion of Persons by the States of Employment and the Unemployment Rate : Rural Maharashtra, NSSO 38-th Round, Combined Sample

States of Employment	Estimated proportion of persons for					All Groups
	ST	SC	Othhind	Othmusm	Othgrp	
(1)	(2)	(3)	(4)	(5)	(6)	
Emp	0.5179	0.4457	0.4613	0.4165	0.4236	0.4622
Unemp	0.0463	0.0547	0.0258	0.0341	0.0463	0.0335
Outlab	0.4358	0.4996	0.5129	0.5494	0.5301	0.5043
Unemprate	0.0820	0.1093	0.0529	0.0757	0.0986	0.0675
Per	3287	2255	15214	1084	1634	23474

2.3.5 Effect of Social Groups

To study the employment pattern of various social groups, the population is divided into five groups. The socioeconomic grouping considered in this dissertation is same as that of Pal (1989). We construct these groups by suitably combining caste and religion. Since in India the scheduled caste and the scheduled tribe households are, in general, economically backward (Pal, 1989), their employment behaviour may be of special interest. We, therefore, form the first two groups as : (i) scheduled tribe (ST) and (ii) scheduled caste (SC). The other three groups are formed according to religion, excluding the SC and ST households from each religious community. These are : (iii) other Hindu (Othhind), (iv) other Muslim (Othmusm) and (v) other households (Othgrp). Estimates relating to employment pattern separately for these social groups are presented in Table 2.3.5.

Table 2.3.5 indicates that the other Hindus enjoy the best position so far as underemployment is concerned. Other Muslims come next to them. However, the percentage of time spent outside labour force is the highest for them. This is perhaps due to *purdah*, a social stricture, which keeps a significant portion of Muslim women consistently outside labour force. This explains why the unemployment rate for the other Muslims is slightly more than that for the other Hindus. The Scheduled Caste and the Scheduled Tribes are obviously the worst affected communities. However, the unem-

Table 2.3.6: Demographic Groupwise Estimated Proportion of Persons by the States of Employment and the Unemployment Rate : Rural Maharashtra, NSSO 38-th Round, Combined Sample

States of Employment	Estimated proportion of persons for						All Groups
	Boychild	Girlchild	Adultmale	Adultfemale	Oldmale	Oldfemale	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Emp	0.1511	0.1180	0.8288	0.4888	0.4841	0.1286	0.4622
Unemp	0.0077	0.0095	0.0580	0.0394	0.0258	0.0104	0.0335
Outlab	0.8412	0.8725	0.1132	0.4718	0.4901	0.8610	0.5043
Unemprate	0.0487	0.0745	0.0654	0.0745	0.0505	0.0749	0.0675
Per	4093	3774	7028	7284	638	657	23474

ployment rate of 'Othgrp' is also very high. This is because in Maharashtra a significant portion of the 'Othgrp' are Neo-Buddhists, who are mostly the converted SC and ST persons.

2.3.6 Effects of Age and Sex

The employment behaviour of individuals are likely to be affected by their demographic particulars. It is well known that employment behaviour of females are very different from that of males. Female employment behaviour is characterised by more frequent withdrawals and entries into the labour market (Bardhan, 1984). Also, as the present data cover all persons with age not less than five completed years, it will be interesting to examine the incidence of child labour. For the purpose of the present study, we have considered six demographic groups of persons, viz., (i) boy children with age below fifteen years (Boychild), (ii) girl children with age below fifteen years (Girlchild), (iii) young adult males with age between fifteen and sixty years (Adultmale), (iv) young adult females, with age between fifteen and sixty years (Adultfemale), (v) old males with age greater than sixty years (Oldmale) and (vi) old females with age greater than sixty years (Oldfemale). The results are presented in Table 2.3.6.

As should be expected, one may notice considerable differences in the pattern of employment behaviour for persons belonging to different age-sex classes. The percentage of time women remain outside labour force is considerably higher than that for men. The female unemployment rate is seen to be slightly higher than that for the male for all the age groups. Such a phenomenon was also noticed by Paul (1988).

Table 2.3.6 reveals a rather unpleasant feature of the Indian labour market – viz., a high incidence of child labour in the rural economy. We find that more than 15% and 11% of the total time of little boys and girls respectively is spent on 'gainful activities'. Although the data reveal that it is the little boy in the family who has to 'work' more than his sister, the actual situation may be quite different. It is possible that the little girl in the family is more often engaged in household chores and thus is counted outside labour force. The extent of sex bias with respect to child employment, can be ascertained by determining the percentage of time spent by boys and girls on different types of activities *when such children are out of labour force* (e.g., time spent on educational activities, time spent on household chores etc.). However, such a detailed study is beyond the scope of this dissertation.

2.3.7 Effect of Education

It is often argued that those who are highly educated compared to the general educational standard of the rural people want to participate in white collar activities only (Visaria and Minhas, 1991). As a consequence, they are often *either* steadily employed *or* steadily unemployed (i.e., waiting for a more 'respectable' job). Even otherwise, the level of education is believed to be a major determinant of an individual's employment behaviour.

To examine the importance of education as an explanatory variable for the present study, we have worked out estimates of various proportions relating to employment behaviour. For this purpose, we have divided the population into three categories, viz., (i) those who are illiterates (Illit), (ii) those who are literates but have not passed the secondary examination (Primid) and (iii) those who have passed the secondary examination (Highed). The results are presented in Table 2.3.7. Since only persons whose ages are greater than or equal to fifteen years can pass the secondary examination and subsequently become highly educated according to our definition, *the various proportions corresponding to the employment states and*

Table 2.3.7: Educational Groupwise Estimated Proportion of Persons by the States of Employment and the Unemployment Rate : Rural Maharashtra, NSSO 38-th Round, Combined Sample

States of Employment	Estimated proportion of persons for			All Groups
	Illit	Primid	Highed	
(1)	(2)	(3)	(4)	(5)
Emp	0.5930	0.6578	0.6590	0.6134
Unemp	0.0484	0.0403	0.0504	0.0488
Outlab	0.3586	0.3019	0.2906	0.3378
Unemprate	0.0755	0.0578	0.0711	0.0798
Per	9603	5754	842	16199

the unemployment rates for this table have been calculated for persons having age greater than or equal to fifteen.

From Table 2.3.7 we see that the rate of unemployment is highest for the illiterates. However, unemployment rate of the educated is close to this. Although the rate of unemployment is somewhat similar for these two categories, their employment patterns may be quite different. The difference in the nature of unemployment can be understood only after we have examined the nature of fluctuation of employment for these two groups in Section 2.4.

2.4 Employment Fluctuation in Rural Maharashtra : A Preliminary Data Analysis

Suppose we observe n individuals for T successive periods. On each period, an individual may be employed (E), unemployed (U) *or* be out of labour force (O). A systematic study of this period-to-period flows of population is absolutely necessary for a better understanding of the labour market. The study of period-to-period movements answers how fast the labour market can approach the steady state from a given state and helps us to identify the segment of the population which is more shock-prone to labour market uncertainties. This kind of exercise may help (i) in forecasting short-run variations in labour market activity (David and Otsuki, 1968) *or* (ii) simulating labour force movements (Denton, 1973).

In this section we shall examine the nature and the extent of employment fluctuation in rural Maharashtra on the basis of the household level information for the State from the NSSO 38-th round enquiry. Traditionally, models on employment fluctuation assume that the individuals' period-to-period employments follow independent Markov chains (David and Otsuki, 1968; Denton, 1973). The different transition probabilities across various subgroups of the population are supposed to reveal the nature and the extent of employment fluctuation. By this method, one can easily compare the elements of the transition matrices pairwise for two different Markov chains, but for comparing the fluctuations of the Markov chains *as a whole* more aggregated measures combining several employment states are needed.

We shall discuss some important aspects of Markov chain models in Chapters 4, 5 and 6. In this chapter, however, we measure employment fluctuation by the total number of state changes ('jumps') in the reference week. For example, suppose an individual's employment vector for the reference week is (E, E, U, E, E, O, E). His number of jumps will be four; one from employment to unemployment, one from unemployment to employment, one from employment to out of labour force and one from out of labour force to employment. In Chapter 5 we shall show that the average number of jumps per day can be interpreted as a *mobility index*, a summary measure of fluctuation.

Here we shall examine the pattern of employment fluctuation for rural Maharashtra as a whole and also separately for groups of individuals belonging to households classified by each of the covariates considered in Section 2.3 above. Needless to mention, the results of this analysis are expected to reveal a comprehensive picture of the labour market in the rural Maharashtra. In what follows, we shall present the cumulative relative frequency distribution (CRFD) of persons by the number of jumps for the entire rural Maharashtra as well as separately for various groups of households classified by the covariates in Section 2.3.

Note that if we treat the number of jumps made by an individual to be a *random variable*, then the CRFD becomes the empirical distribution function (EDF) of the number of jumps.

Let Y_i, $(i = 1, 2, \ldots, n)$ be a random variable which denotes the total number of jumps of the i-th individual in the reference week and let $F(y)$

be its cumulative distribution function (CDF). Then the EDF $F_n(y)$ is

$$F_n(y) = \frac{\#(Y_i \leq y)}{n}, \qquad -\infty < y < \infty \tag{2.4.1}$$

The EDF has been studied extensively for nonparametric data analysis (D'Agostino and Stephens, 1986). It has some definite advantages over other statistical devices, viz.,

(i) The use of EDF plot does not require any assumption regarding the underlying parametric distribution. In fact, it supplies immediate and direct information regarding the shape of the underlying distribution,

(ii) It supplies robust information on location and dispersion. More importantly, for large samples $F_n(y)$ converges uniformly to $F(y)$.

(iii) Its complexity is independent of the number of observations and it does not involve grouping difficulties.

This chapter, however, carries out no theoretical analysis involving the EDF of jumps. Comparison of the EDF with a proposed CDF will be done in Chapter 3. The EDF for different covariate categories will only be presented here. For convenience, the average (Avg) and the variance (Var) of the distribution of the number of jumps are reported along with the number of sample persons (Per) in the tables presented here.

2.4.1 Employment Fluctuation in the Aggregate

Table 2.4.1 presents the estimated distribution of the number of jumps made by individuals in the reference week for rural Maharashtra as a whole. For this distribution the average number of jumps works out to be 0.2102 which implies that roughly 3% of individuals change state on an average per day. The variance of the distribution is 0.4055. Given that Table 2.4.1 is calculated including women and children (a significant percentage of whom remain consistently out of labour force), the average should still be considered very high by any standards.

It therefore appears that the extent of employment fluctuation for the State as a whole is not negligible, specifically in view of the fact that the EGS was continually in operation in the State which should have a stabilising effect on the rural labour market.

Table 2.4.1: Cumulative Relative Frequency Distribution (CRFD) of Jumps : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Number of Jumps	0	1	2	3	4	5	6
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CRFD	0.8780	0.9347	0.9846	0.9936	0.9990	0.9999	1.0000

Avg= 0.2102, Var=0.4055, Pcr=23474

2.4.2 Employment Fluctuation across Regions

Table 2.4.2 should help us to understand the regional patterns of employment fluctuation. It shows that the Coastal area of Maharashtra has the most stable employment pattern. The cumulative relative frequency of jump for this region lies uniformly above all other regions. The average of jumps and the variance of jump for this region are also the smallest. Note that the unemployment rate of Coastal Maharashtra is 5.78%, which is slightly higher than the corresponding figure of 4.89% of the more developed Inland Western region. However, the Coastal region of Maharashtra enjoys a steadier amount of adequate rainfall than the more drought prone Inland Western region, resulting in a steadier agricultural activities in that region. Possibly that is why the number of jumps made by individuals in Coastal Maharashtra is so low.

It may be noted that the distributions for the Inland Western, Inland Central and the Eastern regions are somewhat similar in nature. We have earlier seen that the rate of unemployment for these three regions are widely different, the rates being 4.89%, 5.82% and 9.07% respectively. In fact, the lower unemployment rate for the Inland Western region may be due to the economically more developed nature of the region. The high unemployment rate of the Eastern region, on the other hand, may be due partly to its economic backwardness. However, the spread of irrigational infrastructure in this region may have had a stabilising effect on employment.

Table 2.4.2 also identifies the Inland Northern and the Inland Eastern

Table 2.4.2: Regionwise Cumulative Relative Frequency Distribution of Jumps : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Number of Jumps	CRFD for the region						All Regions
	Coastal	Western	Northern	Central	Incast	Eastern	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	0.9431	0.8948	0.8281	0.8946	0.8269	0.8752	0.8780
1	0.9687	0.9416	0.9010	0.9473	0.9128	0.9329	0.9347
2	0.9951	0.9856	0.9751	0.9857	0.9790	0.9906	0.9846
3	0.9972	0.9945	0.9884	0.9944	0.9928	0.9943	0.9936
4	1.0000	0.9997	0.9978	0.9982	0.9985	0.9994	0.9990
5	1.0000	1.0000	1.0000	0.9998	1.0000	0.9994	0.9999
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Avg	0.0959	0.1837	0.3096	0.1800	0.2899	0.2070	0.2102
Var	0.1858	0.3596	0.5988	0.3604	0.5195	0.3782	0.4055
Per	2495	6793	3953	3976	4266	1991	23474

parts of Maharashtra as having more employment fluctuation compared to the other regions. The average number of jumps in these parts are more than thrice that of the Coastal area. The reason for this high degree of employment fluctuation in these regions is possibly due to the combined effect of their economic backwardness and lack of an adequate irrigational infrastructure.

2.4.3 Employment Fluctuation across Subrounds

We have already seen that the rate of unemployment in rural Maharashtra shows some variations over the subrounds of the NSSO enquiry. The source of this seasonal variation is possibly the seasonality of agricultural operations with which the rural labour market is closely connected. It would be interesting to examine if such seasonality affects the pattern of employment fluctuation as well. Table 2.4.3 presents the subroundwise estimates of the CRFD for the number of jumps for the rural population in Maharashtra.

Table 2.4.3 reveals that the patterns of employment fluctuation for all the four subrounds are broadly similar. The relative cumulative frequency of the number of jumps for Subround 3 (July – September) is below those for

Table 2.4.3: Subroundwise Cumulative Relative Frequency Distribution of Jumps : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Number of Jumps	CRFD for the Subround				All Subrounds
	Subround 1 (Jan-Mar)	Subround 2 (Apr-Jun)	Subround 3 (Jul-Sep)	Subround 4 (Oct-Dec)	
(1)	(2)	(3)	(4)	(5)	(6)
0	0.8934	0.8827	0.8649	0.8684	0.8780
1	0.9410	0.9381	0.9252	0.9327	0.9347
2	0.9888	0.9859	0.9738	0.9878	0.9846
3	0.9960	0.9934	0.9896	0.9948	0.9936
4	0.9998	0.9978	0.9974	0.9997	0.9990
5	1.0000	0.9997	1.0000	1.0000	0.9999
6	1.0000	1.0000	1.0000	1.0000	1.0000
Avg	0.1810	0.2025	0.2480	0.2167	0.2102
Var	0.3365	0.4025	0.5159	0.3873	0.4045
Per	6162	6203	5032	6077	23474

the other three subrounds (which means that the average number of jumps for Subround 3 is the highest). This seems to support the idea that during the 'peak' season employment fluctuates more. However, a more detailed and careful study of seasonality in the pattern of employment fluctuation is required to obtain more definite results. More importantly, perhaps, one needs to examine the interaction effects of pairs of variables as subround and region, subround and occupational type of households etc. so as to bring out the more subtle influences of season on the nature and the degree of employment fluctuation.

2.4.4 Employment Fluctuation across Households' Occupational Type

We have already seen in Section 2.3 that the unemployment rate varies widely across occupational groups in rural Maharashtra. The wage dependent households were seen to be the ones most affected by unemployment as these households did not have an adequate asset base to protect them against labour market shocks. Majority of these wage dependent households being poor, they cannot afford to remain unemployed for a long time. Therefore, they tend to grab any available work opportunity, work for a

Table 2.4.4: Occupationwise Cumulative Relative Frequency Distribution of Jumps : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Number of Jumps	CRFD for the occupational type					All Households
	Selfagr	Selfnagr	Agrlab	Othlab	Othocc	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0	0.9390	0.9370	0.7851	0.8421	0.9553	0.8780
1	0.9689	0.9674	0.8866	0.8957	0.9758	0.9347
2	0.9937	0.9919	0.9704	0.9870	0.9924	0.9846
3	0.9978	0.9953	0.9873	0.9964	0.9965	0.9936
4	0.9998	0.9992	0.9975	1.0000	1.0000	0.9990
5	1.0000	1.0000	0.9998	1.0000	1.0000	0.9999
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Avg	0.1007	0.1091	0.3733	0.2788	0.0800	0.2102
Var	0.1925	0.2290	0.6774	0.4834	0.1737	0.4055
Per	10106	1926	8475	1466	1501	23474

while and once the period of contract is over, become unemployed again. It would be interesting to compare the extent of employment fluctuation of these households with the rest of the population.

Table 2.4.4 presents the CRFD for the number of jumps of such different occupational categories. It reveals that the cumulative relative frequency of jump for the agricultural labourers lies below that of all other categories. Both the average and the variance of the number of jumps for this category are higher than those of the other categories. Average number of jumps made by these households per day is more than 0.05, implying that on average more than one out of twenty persons change state on each day. Moreover, since the variance is also high, a change in the labour market situation may significantly increase or decrease this number.

As regards employment fluctuation, other labour households perform slightly better. Although these households have steadier employment compared to the agricultural labour households, the fluctuation of employment for these households is considerably more than that of the other three occupational categories.

The distributions of jumps for the self-employed in agriculture and the self-employed in non-agriculture look remarkably alike. Although the aver-

Table 2.4.5: Social Groupwise Cumulative Relative Frequency Distribution of Jumps : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Number of Jumps	CRFD for the social group					All Categories
	ST	SC	Othhind	Othmusm	Othgrp	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0	0.8186	0.8320	0.9017	0.8902	0.8336	0.8780
1	0.8953	0.9070	0.9483	0.9405	0.9220	0.9347
2	0.9776	0.9717	0.9884	0.9831	0.9822	0.9846
3	0.9879	0.9866	0.9960	0.9937	0.9932	0.9936
4	0.9984	0.9976	0.9995	0.9984	0.9978	0.9990
5	0.9996	1.0000	0.9999	1.0000	1.0000	0.9999
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Avg	0.3226	0.3051	0.1662	0.1940	0.2711	0.2102
Var	0.6068	0.6109	0.3173	0.3934	0.4828	0.4055
Per	3287	2255	15214	1084	1634	23474

age number of jumps for the non-agricultural self-employed is slightly more than that of the agricultural self-employed, the difference is not much.

2.4.5 Employment Fluctuation across Social Groups

We have already seen that the rate of unemployment in rural Maharashtra shows some variation across caste and religion. We now examine the extent of employment fluctuation across different social and religious communities. The classification constructed for this purpose is the same as in Subsection 2.3.5.

In Table 2.4.5 the distributions of jumps for the different social groups have been compared. Consistent with our a priori expectation, the ST households are seen to be more prone to labour market fluctuations. They are closely followed by the SC and Othgrp households. The distributions of jumps for these three subgroups dominate those for the Othhind and the Othmusm. Note that the CRFD for the Othhind is very similar to that for the Othmusm implying that the pattern of employment fluctuation is similar for the two communities.

Table 2.4.6: Demographic Groupwise Cumulative Relative Frequency Distribution of Jumps : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Number of Jumps	CRFD for the demographic group						All Categories
	Boychild	Girlchild	Adultmale	Adultfemale	Oldmale	Oldfemale	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	0.9807	0.9527	0.8253	0.8223	0.9136	0.9587	0.8750
1	0.9898	0.9749	0.9058	0.9053	0.9601	0.9745	0.9347
2	0.9981	0.9939	0.9810	0.9747	0.9869	0.9938	0.9846
3	0.9996	0.9982	0.9926	0.9887	0.9934	0.9963	0.9936
4	0.9998	0.9997	0.9986	0.9985	0.9982	0.9982	0.9990
5	1.0000	1.0000	1.0000	0.9999	0.9982	1.0000	0.9999
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Avg	0.0320	0.0805	0.2966	0.3106	0.1495	0.0785	0.2102
Var	0.0631	0.1616	0.5282	0.5855	0.3307	0.1848	0.4045
Per	4093	3774	7028	7284	638	657	23474

2.4.6 Employment Fluctuation across Age and Sex

The nature of employment pattern is usually quite different for the demographic groups. For example, females have more frequent entries and withdrawals compared to males (Bardhan, 1984). It would be interesting to examine whether any such systematic difference is present in the patterns of employment fluctuation of the various demographic groups. Table 2.4.6 presents the cumulative relative frequency of jumps for the six demographic categories that we have considered in the present study.

As Table 2.4.6 shows, young women do seem to have greater employment fluctuations than other demographic groups. So far as the children are concerned, the little girls seem to move more frequently across the employment states, as the distribution of jumps for girls tends to dominate that for the boys. However, this difference is observed to be less sharp between adult males and adult females. In fact, for the age group sixteen to sixty years, the distribution of jumps for males and females are remarkably similar and the average and the variance of the number of jumps of the two sexes are quite close. However, though the two distributions look quite similar, one should not conclude that the nature of employment fluctuation is similar for the two sexes because women usually move more across the two states employ-

Table 2.4.7: Educational Groupwise Cumulative Relative Frequency Distribution of Jumps : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Number of Jumps	CRFD for the educational category			All Categories
	Illit	Primid	Highed	
(1)	(2)	(3)	(4)	(5)
0	0.8084	0.8621	0.9500	0.8348
1	0.8991	0.9223	0.9763	0.9113
2	0.9754	0.9833	0.9977	0.9793
3	0.9894	0.9929	1.0000	0.9911
4	0.9983	0.9989	1.0000	0.9985
5	0.9999	0.9998	1.0000	0.9998
6	1.0000	1.0000	1.0000	1.0000
Avg	0.3295	0.2407	0.0761	0.2847
Var	0.5995	0.4580	0.1270	0.5285
Per	9603	5754	842	16199

ment and out of labour force, whereas men are more likely to move between employment and unemployment.[7] Finally, the employment fluctuation of old women is seen to be considerably less than that of old men.

2.4.7 Employment Fluctuation across Education

In Section 2.3, the unemployment rate across different educational categories has been examined. However, such an analysis is far from adequate as the *duration of unemployment* may vary across categories of individuals. For example, educated persons may experience prolonged unemployment due to high reservation wage (Visaria and Minhas, 1991). An analysis of employment fluctuation should reveal whether such a hypothesis is true. In Table 2.4.7 we present the cumulative relative frequency distribution of jumps for different educational categories.

[7]Bardhan(1978) has explored the phenomenon of discouraged dropout of women in the context of Indian rural labour market, which increases the movement between unemployment and out of labour force. Such movements can be observed in long-run data. Since we are investigating the labour market behaviour of individuals for seven consecutive days only, the occurrence of discouraged dropout will not be much.

From Table 2.4.7, it becomes quite clear that the level of education affects the pattern of employment fluctuation. The nature of employment fluctuation of the highly educated is quite different from the rest of the population. The CRFD of the number of jumps for the highly educated lies uniformly above than that for the two other categories. The average jumps per day of the highly educated is slightly more than 0.01, implying that only one out of hundred persons in this category changes his/her respective employment state per day.

2.5 Conclusion

A descriptive data analysis on the nature and the extent of underemployment and employment fluctuation in rural Maharashtra is presented in this chapter. The analysis has been carried out at the aggregate level as well as separately for some important covariates. Our results indicate that in spite of the successful operation of the EGS in Maharashtra, the State had considerable underemployment. It is also found that there is considerable heterogeneity in employment patterns across various subgroups of the population.

The empirical analysis reported here clearly indicates some covariates across which employment patterns vary widely. However, one should be cautious while drawing any firm conclusion on the basis of the results presented here because of the following limitations of the study. First, in the NSSO employment and unemployment enquiry, information is collected from sample households for a very brief reference period, viz., one week. The extent of underemployment and employment fluctuation reported here may get affected by the disorders prevailing in the labour market during the brief reference period. Since labour market adjustments may take some time, once a distortion from the steady state occurs, daily status measures throughout the reference week is likely to be affected.

Second, here we have considered one covariate at a time in our analysis. This is unlikely to give the true partial effect of each covariate in the presence of significant interaction effect. In subsequent chapters, we shall attempt to examine such interactions by specifying appropriate analytical models.

Chapter 3

Employment Fluctuation in Rural India : A Poisson Regression Model

3.1 Introduction

In Chapter 2, we examined the extent and the nature of employment fluctuation separately across various covariates. Although such a study might provide some important insight into the nature of the problem, the true partial effect of each covariate might be, to some extent, lost due to aggregation. In this chapter, we propose a general model which accounts for the variation in employment status measured in terms of the number of jumps of individuals *simultaneously* across covariates.

For an individual, the total number of jumps in the reference period is an integer variable. If the reference period is short, it will take only a few integer values and hence should not be approximated by a continuous variable. Since the dependent variable becomes a *count* variable, application of ordinary least squares (OLS) will be inappropriate and a suitable model will be needed to deal with the problems arising due to the specific nature of the dependent variable.

In this chapter, we shall use Poisson regression technique which will be an appropriate tool in the present case. Poisson regression models have recently received much attention in econometric literature as models for analysing count data. These models have been used to explain firms' patents (Hausman *et al.*,1984), the number of doctor consultations made (Cameron and Trivedi, 1986), daily beverage consumption (Mullahy,1986), daily homicide counts (Grogger, 1990), number of malpractice claims (Cooil,1991) and number of books ordered (Wedel *et al.*,1993). Poisson regression models have a number of attractive features. They accommodate the integer property of the count data directly and help justify aggregation of the count variable over time. However, the Poisson regression model involves the restrictive assumption that the variance is equal to the mean. In many economic applications the variance of the *count* variable exceeds its mean. This situation is known as the problem of 'overdispersion' (Cameron and Trivedi, 1986). This problem is resolved by assuming that the Poisson parameter λ varies randomly across the population with a certain probability distribution. It is generally assumed that λ follows a Gamma distribution, as this assumption yields an easily interpretable closed form solution. The resulting compound distribution is Negative Binomial and is used frequently in the literature (Hausman *et al.*, 1984 ; Cameron and Trivedi, 1986).

The plan of this chapter is as follows : Section 3.2 presents the basic analytical framework. Section 3.3 presents the empirical results and finally Section 3.4 summarises the main findings.

3.2 The Analytical Framework

Suppose there are n individuals and the employment position of each individual is being observed for $(T+1)$ discrete points of time at equal time interval.[1] For convenience, we shall call each such interval a *day.* The actual starting point of observation may vary from individual to individual, i.e., the reference period may be moving.

On any given day, a person may be in any of the three states, viz., (i) employed (E), (ii) unemployed (U) *or* (iii) out of labour force (O). Suppose at the beginning, i.e., on day 0, the i-th individual is in state x_{0i}, where x_{0i} can be either E, U or O. Then on day τ_{1i}, he jumps to x_{1i}, $x_{1i} \neq x_{0i}$. and after some days, he again jumps to a state x_{2i}, $x_{2i} \neq x_{1i}$ on day τ_{2i}, $\tau_{2i} > \tau_{1i}$.

[1]Note that NSSO observes all households for an equal number of days. We can, however, easily generalise the model to unequal time intervals across individuals.

Such movements from one state to another take place indefinitely – though we may observe the system only upto day T. We assume that none of the states E, U or O is an *absorbing state*, i.e., movement from each state to some other state is possible.

Now let $X_i(t), i = 1, 2, \ldots, n;\ t = 0, 1, \ldots, T$ denote the state of the i-th individual on day t. We assume that $X_i(t)$'s are independent across i. Here,

$$X_i(t) = \begin{cases} x_{0i}, & 0 \leq t < \tau_{1i} \\ x_{1i}, & \tau_{1i} \leq t < \tau_{2i},\ x_{1i} \neq x_{0i} \\ x_{2i}, & \tau_{2i} \leq t < \tau_{3i},\ x_{2i} \neq x_{1i} \\ \cdots & \cdots \end{cases} \tag{3.2.1}$$

Corresponding to each $X_i(t)$, we now define another related stochastic process $Y_i(t)$ as follows

$$Y_i(t) = \begin{cases} 0, & 0 \leq t < \tau_{1i} \\ 1, & \tau_{1i} \leq t < \tau_{2i} \\ 2, & \tau_{2i} \leq t < \tau_{3i} \\ \cdots & \cdots \end{cases} \tag{3.2.2}$$

The process $Y_i(t)$ gives the total number of jumps observed for the i-th individual till day t. For finite T, the state space of the process is

$$S = \{0, 1, \ldots, T\} \tag{3.2.3}$$

Note that since $X_i(t)$'s are independent across i, so are the $Y_i(t)$'s.

We shall assume that

$$Y_i(T) \sim \text{ Poisson}(\lambda_i) \tag{3.2.4}$$

Let $f(y_i)$ denote the probability that the i-th individual will jump y_i times in the reference period. Here,

$$f(y_i) = P[Y_i(T) = y_i] = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{3.2.5}$$

We shall now specify how this parameter λ_i varies across individuals, viz.,

$$\lambda_i = e^{\widetilde{\mathbf{Z}}_{\mathbf{i}}'\widetilde{\beta}} \tag{3.2.6}$$

where $\widetilde{\mathbf{Z}}_{\mathbf{i}}$ is a $K \times 1$ vector of characteristics of the i-th individual and $\widetilde{\beta}$ is the corresponding parameter vector. This specification of λ_i has some special advantages. It is analogous to the familiar regression specification. Moreover, this ensures that the estimated value of λ_i is always nonnegative, thus ensuring meaningful results. Given (3.2.5) and (3.2.6) the log-likelihood function for the n individuals is

$$\mathcal{L}(\widetilde{\beta}) = \sum_{i=1}^{n}[-\log_e(y_i!) - e^{\widetilde{\mathbf{Z}}_i'\widetilde{\beta}} + y_i\widetilde{\mathbf{Z}}_i'\widetilde{\beta}] \tag{3.2.7}$$

The gradient and the Hessian can be written as

$$\frac{\partial \mathcal{L}}{\partial \widetilde{\beta}} = \sum_{i=1}^{n}[\widetilde{\mathbf{Z}}_i'(y_i - e^{\widetilde{\mathbf{Z}}_i'\widetilde{\beta}})] \tag{3.2.8}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \widetilde{\beta}\partial \widetilde{\beta}'} = -\sum_{i=1}^{n}[\widetilde{\mathbf{Z}}_i\widetilde{\mathbf{Z}}_i' e^{\widetilde{\mathbf{Z}}_i'\widetilde{\beta}}] \tag{3.2.9}$$

The regression property of this specification comes from the fact that $\mathbf{E}[Y_i(T)] = \lambda_i$, so that $(y_i - \lambda_i)$ may be interpreted as the 'residual' and the parameter vector $\widetilde{\beta}$ can be estimated by an iterative nonlinear weighted least squares method. Alternatively, the method of maximising the likelihood function may also be used. Under some mild conditions, the likelihood function becomes globally concave and convergence takes place rapidly (Hausman *et al.*, 1984).

Even this general framework fails to tackle the problem of overdispersion. However, the problem may be solved by introducing another source of randomness. We allow the Poisson parameter λ_i to be randomly distributed across the population and assume that λ_i follows a Gamma distribution,

$$\lambda_i \sim \Gamma(\gamma_i, \delta) \tag{3.2.10}$$

where

$$\gamma_i = e^{\widetilde{\mathbf{Z}}_i'\widetilde{\beta}} \tag{3.2.11}$$

so that the mean and the variance of λ_i are, $\mathbf{E}(\lambda_i) = \frac{\gamma_i}{\delta}$ and $\mathrm{Var}(\lambda_i) = \frac{\gamma_i}{\delta^2}$, respectively.

The compound distribution of $Y_i(T)$ is now given by

$$g(y_i) = \frac{\Gamma(\gamma_i + y_i)}{\Gamma(\gamma_i)\Gamma(y_i+1)}\left(\frac{\delta}{1+\delta}\right)^{\gamma_i}(1+\delta)^{-y_i} \tag{3.2.12}$$

which is a Negative Binomial distribution with parameters (γ_i, δ). The first two moments of $Y_i(T)$ are given by, $\mathbf{E}[Y_i(T)] = \frac{\gamma_i}{\delta}$ and $\mathrm{Var}[Y_i(T)] = \frac{\gamma_i(1+\delta)}{\delta^2}$.

Therefore, the variance to mean ratio is

$$\frac{\mathrm{Var}[Y_i(T)]}{\mathrm{E}[Y_i(T)]} = \frac{1+\delta}{\delta} > 1 \tag{3.2.13}$$

which should take care of the phenomenon of overdispersion mentioned above. Note that the limiting case of Poisson distribution is reached when $\delta \to \infty$.

3.3 Empirical Results

In this section we shall present the empirical results of the Poisson and Negative Binomial regression analysis that have been obtained for the rural households of the state of Maharashtra on the basis of the NSSO 38-th round survey data on employment and unemployment for the State. Subsection 3.3.1 presents a brief discussion on the choice of covariates. In Subsection 3.3.2, we specify the set of variables to be included in Poisson and Negative Binomial regression. In Subsection 3.3.3, we report the results of our estimation and finally in Subsection 3.3.4, we examine the goodness of fit of the models proposed.

3.3.1 Choice of Covariates

The covariates chosen for the empirical study are the same as those used in Chapter 2, viz., (i) geographical area (region), (ii) agricultural season (subround), (iii) household's main occupational type, (iv) household's social and religious group, and (v) age, (vi) sex and (vii) education of individuals. Each of these covariates has several categories. However, for computational tractability (e.g., the convergence of the maximum likelihood iterative procedure) we have suitably regrouped these categories to reduce the number of parameters to be estimated. For a preliminary analysis of the available data on 'jumps' across employment states, in Chapter 2 we worked out the cumulative frequency distribution of the number of jumps separately for each category of each of the covariates. We also calculated some descriptive summary statistics (like average number of jumps, variance of jumps etc.) categorywise for each covariate. We again present some of these summary measures from Chapter 2 along with some more such measures in

Table 3.3.1. These measures will be needed to examine the extent of overdispersion across categories and also will help us in specifying an appropriate model.

3.3.2 Choice of Explanatory Variables

Table 3.3.1 confirms the existence of wide heterogeneity across various regions in Maharashtra (Dev, 1992). It is well known that the neighbourhood of Bombay (which falls in the Coastal region), and the Western part of Maharashtra are highly developed. The rest of the state is grossly underdeveloped relative to these two regions. The average number of jumps for Coastal which is 0.0959 and is much smaller compared to those for the other regions is perhaps due to this reason. For the statistical analysis that follows we have clubbed the Coastal and the Inland Western regions of Maharashtra by specifying a dummy variable WEST which takes the value unity if an individual belongs either to the Coastal or to the Inland Western region and takes the value zero otherwise.

The summary statistics on the number of jumps for different subrounds do not show wide variation. The average number of jumps is little higher for Subrounds 3 and 4 compared to those for Subrounds 1 and 2. A priori, one would normally expect to see more dramatic variations in the result across subround because of inherent seasonality in agricultural operations which is the main source of rural employment in India. However, it is quite likely that such seasonality in individual employment fluctuations can only be seen when one considers the interaction between season and other covariates. We shall, therefore, estimate four separate models for the four subrounds.

Numerous studies have indicated that the employment pattern of the self-employed is markedly different from that of labourers (Visaria and Minhas, 1991). The results in Table 3.3.1 are fully consistent with these previous findings. Here, the average number of jumps observed for the persons belonging to the labour households is around three times that for their self-employed counterparts. In subsequent analysis, this difference in employment patterns of the self-employed and the labourers is captured by introducing a dummy variable called LABOUR which is assigned a value of unity for persons belonging to agricultural labour households and other labour households and zero otherwise.

Table 3.3.1: Descriptive Statistics of Jumps Separately by the Subgroups of the Covariates : Rural Maharashtra, 38-th Round, Combined Sample

Covariates	Categories	Average (A)	Variance (V)	V/A	C.V.[1]	Sample
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Region	Coastal	0.0959	0.1858	1.9374	449.47	2495
	Western	0.1837	0.3596	1.9575	326.44	6793
	Northern	0.3096	0.5988	1.9341	249.94	3953
	Central	0.1800	0.3604	2.0022	333.52	3976
	Ineast	0.2899	0.5195	1.7920	248.62	4266
	Eastern	0.2070	0.3782	1.8271	297.09	1991
Subround	Subround 1	0.1810	0.3365	1.8591	320.49	6162
	Subround 2	0.2025	0.4025	1.9876	313.30	6203
	Subround 3	0.2480	0.5159	2.0802	289.62	5032
	Subround 4	0.2167	0.3873	1.7873	287.19	6077
Household Type	Selfagr	0.1007	0.1925	1.9116	435.70	10106
	Selfnagr	0.1091	0.2290	2.0990	438.62	1926
	Agrlab	0.3733	0.6774	1.8146	220.48	8475
	Othlab	0.2788	0.4834	1.7339	249.38	1466
	Othocc	0.0800	0.1737	2.1712	520.97	1501
Social Group	ST	0.3226	0.6068	1.8810	241.47	3287
	SC	0.3051	0.6109	2.0023	256.18	2255
	Othhind	0.1662	0.3173	1.9091	338.92	15214
	Othmusm	0.1940	0.3934	2.0278	323.31	1084
	Othgrp	0.2711	0.4828	1.7809	256.30	1634
Age and Sex	Boychild	0.0320	0.0631	1.9719	784.99	4093
	Girlchild	0.0805	0.1616	2.0074	499.37	3774
	Adultmale	0.2966	0.5282	1.7808	245.03	7028
	Adultfemale	0.3106	0.5855	1.8851	246.36	7284
	Oldmale	0.1495	0.3307	2.2120	384.66	638
	Oldfemale	0.0785	0.1848	2.3541	547.62	657
Education[2]	Illit	0.3295	0.5995	1.8194	234.98	9603
	Primid	0.2407	0.4580	1.9027	281.16	5754
	Highed	0.0761	0.1270	1.6689	468.29	842

[1] C.V. denotes coefficient of variation, C.V.=$(\sqrt{V}/A) \times 100$

[2] For Education, measures are calculated for those having age greater than fifteen completed years.

So far as the effect of social class is concerned. we find that the problem of employment fluctuation is more serious for the scheduled caste (SC) and the scheduled tribes (ST). The average number of jumps observed for them is almost twice that for other Hindus (Othhind) or other Muslims (Othmusm). The remaining social class, viz., Othgrp (i.e., the one consisting of Neo-Buddhists among others) also has a high average, possibly because of the fact that many of the households in this class are converted SC/ST. To distinguish the persons belonging to SC and ST households from others in the analysis that follows, we used a dummy variable SCST which was assigned the value unity for persons who were either scheduled caste or scheduled tribe. This was done especially because the variable SCST has an important sociopolitical implication in the Indian context.

To examine the effect of age and sex on employment fluctuation, we considered in the preliminary data analysis the same six age-sex categories specified in Chapter 2. Table 3.3.1 suggests that both age and sex considerably affect employment fluctuation. Moreover, the effect of age on average number of jumps appear to be nonlinear, because, for either sex the average is seen to be higher for the adult category compared to that for the child and old categories. In order to capture the effect of age and sex in the present analysis, we have defined three dummy variables. The dummy variable FEMALE (which takes the value of unity for a female and zero otherwise) is introduced to capture the differential effect of sex and two other dummy variables, viz., CHILD and OLD are introduced to capture the nonlinear effect of age. It may be noted that here we have defined a child as a person of age between five and fourteen years and by an old person we mean a person having age greater than sixty years.

It is quite well known that in the rural sector the employment pattern of the highly educated are markedly different from other persons (Visaria and Minhas, 1991). If they are employed, they are generally employed in some white-collar jobs and continue to remain employed. However, among the educated unemployed persons, some may have high reservation wage and wait for better jobs and hence remain consistently unemployed. Also, some of the educated persons are engaged in education and hence are counted as consistently out of labour force in the reference period. This means, their employment status displays more stability than other educational subgroups of the population. Our results in Table 3.3.1 support this view. The average number of jumps observed for the highly educated (Highed) is only 0.0761 while for the illiterates (Illit) it is as high as 0.3295. Those who studied upto secondary level (Primid) fall somewhere in between, the average for them being 0.2407. In view of this noticeable difference in the result for the highly educated person, we define an education dummy HIGHED which takes the value one if the person concerned has at least secondary education and zero otherwise.

Besides the extent of variation of jumps across covariates, Table 3.3.1 also displays the presence of the phenomenon of overdispersion. For all subgroups corresponding to all covariates, the problem of overdispersion exists. The ratio of variance to mean generally varies between 1.7 and 2.1, indicating that a Negative Binomial model, with the overdispersion parameter δ being approximately equal to unity, would perhaps fit the observations.

3.3.3 Model Choice and Estimation

As we have already mentioned, the distribution of jumps and the descriptive statistics for the same were found to be more or less similar for all the four subrounds. However, we decided to estimate the models separately for different subrounds. This was done essentially to capture the interaction effect, if any, between seasonality and other possible explanatory factors of employment fluctuation.

The Poisson and the Negative Binomial regression models were estimated following the maximum likelihood method using the computer software package SHAZAM (Version 6.2 by White, 1990). Because of the parametric nonlinearities of the estimated model, an iterative estimation procedure was adopted. A maximum absolute difference of 0.00001 of all the coefficients in successive iteration was taken as the rule of convergence. Given this rule, the number of iterations required for convergence varied between 15 and 25 in most of the cases.

We present the estimates of the Poisson and Negative Binomial model in Table 3.3.2. The estimated standard errors of the coefficients are also presented below each coefficient in brackets. The standard errors are calculated using the estimates obtained separately from the subsamples. For any parameter θ, if the estimates from the subsamples are $\widehat{\theta_1}$ and $\widehat{\theta_2}$ respectively, the estimated standard error s_θ is calculated as[2]

$$s_\theta \doteq \frac{1}{2} \mid \widehat{\theta}_1 - \widehat{\theta}_2 \mid \qquad (3.3.1)$$

This method provides an easy way to estimate the standard errors. However, s_θ has only one degree of freedom and therefore may not be very reliable from the point of view of statistical inference. Note that the observed t-ratio,

$$t_o = \frac{\widehat{\theta}}{s_\theta} \qquad (3.3.2)$$

follows t-distribution with only one degree of freedom. Therefore, for two tailed tests, a coefficient in the model will be significantly different from zero at 5% and 1% level if it is larger than 12.71 and 63.66 respectively.

Table 3.3.2 reveals that for all the four subrounds, the Negative Binomial model provides a better fit to the given data in terms of the log-likelihood values. Even though the estimated value of the common parameters (i.e., all parameters except δ) are quite close for all the four subrounds, the values of log-likelihood for the Negative Binomial models are much higher than those corresponding to the Poisson model.

Table 3.3.2 also reveals that the value of the overdispersion parameter δ varies between 0.79 and 1.09, which is consistent with our previous finding that for almost all covariate groups, sample variance is approximately double that of the sample mean. Also, in all cases the combined sample estimates of the parameters are found to fall between the two corresponding estimates from the subsamples.[3]

Table 3.3.2 reaffirms the fact that the labourers are far more prone to employment fluctuation than their self-employed counterparts. The estimated coefficients of the dummy variable LABOUR vary between 1.0 and 1.35 for the Negative Binomial model. This means, if the other covariates remain same, a labourer on an average jumps 3 times more than a self-employed person. These estimated coefficients also display considerable robustness across subrounds, though the estimate for Subround 2, i.e., for April to June, is somewhat smaller.

For all the subrounds, children and old persons are found to have stable employment patterns than their young adult counterparts. This is not

[2]See Appendix A for details.

[3]The results for the subsamples are not reported here.

Table 3.3.2: The Estimates of the Poisson and the Negative Binomial Models of Employment Fluctuation : Rural Maharashtra, NSSO 38-th Round, Combined Sample

	Subround 1		Subround 2		Subround 3		Subround 4	
Coefficients	Poisson	Neg Bin	Poisson	Neg Bin	Poisson	Neg Bin	Poisson	Neg Bin
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CONSTANT	$-1.8634^{\ddagger}$	$-1.9520^{\ddagger}$	-1.5878	-1.7341^{**}	$-1.8414^{\ddagger}$	$-2.1216^{\ddagger}$	$-2.0956^{\ddagger}$	$-2.0119^{\ddagger}$
	$(0.0133)^{*}$	(0.0397)	(0.2697)	(0.2187)	(0.0406)	(0.0852)	(0.0624)	(0.0929)
WEST	-0.7066	-0.7670	-0.1962	-0.1920	-0.3372	-0.3571	-0.0923	-0.1813
	(0.2456)	(0.1967)	(0.0595)	(0.0507)	(0.1202)	(0.1729)	(0.1213)	(0.1591)
LABOUR	$1.2760^{\ddagger}$	$1.3414^{\ddagger}$	0.9022	0.9929^{**}	$1.2271^{\ddagger}$	$1.3499^{\ddagger}$	$1.2333^{\ddagger}$	$1.2926^{\ddagger}$
	(0.0194)	(0.0928)	(0.1662)	(0.1079)	(0.0245)	(0.0017)	(0.0304)	(0.0720)
SCST	0.0413	0.0926	$0.3790^{\ddagger}$	0.2866^{**}	0.3083	0.2691	0.2107	0.2311
	(0.2076)	(0.1814)	(0.0223)	(0.0277)	(0.2203)	(0.1862)	(0.1420)	(0.1650)
CHILD	-1.7298	-1.8425	$-2.0038^{\ddagger}$	$-2.1346^{\ddagger}$	$-1.6102^{\ddagger}$	-1.6697^{**}	$-1.6353^{\ddagger}$	$-1.7637^{\ddagger}$
	(0.3065)	(0.3056)	(0.0689)	(0.0298)	(0.1099)	(0.1621)	(0.0366)	(0.0116)
OLD	$-0.8884^{\ddagger}$	$-0.7583^{\ddagger}$	-0.5964^{**}	$-0.6211^{\ddagger}$	-1.1874^{**}	-1.6302^{**}	-0.9243	-1.1472
	(0.0403)	(0.0017)	(0.0860)	(0.0213)	(0.1690)	(0.2283)	(0.4647)	(0.4267)
FEMALE	0.0917	0.0720	-0.3604	-0.4243	0.2852	0.2716	0.3119	0.3150
	(0.0433)	(0.0309)	(0.2161)	(0.2190)	(0.0703)	(0.1257)	(0.0983)	(0.0796)
HIGHED	-0.6426	-0.5321	$-1.2332^{\ddagger}$	$-1.2575^{\ddagger}$	$-0.9132^{\ddagger}$	-1.0963^{**}	-0.9450	-0.7956
	(0.1892)	(0.1480)	(0.0739)	(0.0192)	(0.0608)	(0.0967)	(0.3900)	(0.4842)
δ		0.9510^{**}		$0.8606^{\ddagger}$		$0.7926^{\ddagger}$		1.0948^{**}
		(0.1091)		(0.0095)		(0.0237)		(0.1243)
Log-likelihood	-2856.16	-2529.51	-3206.90	-2799.66	-2946.60	-2544.17	-3228.62	-2915.75
Sample Size	6162		6203		5032		6077	

* The bracketed figures are estimated standard errors

** Significant at 5% level, one sided test

‡ Significant at 5% level, two sided test

surprising as children and old persons generally remain out of labour force. Similarly, the highly educated are seen to have fewer jumps. This is probably because those who are highly educated wait for a 'good' job and once they get it, they remain employed for a long time. Thus they are either steadily employed or steadily unemployed.

The pattern of employment across these covariates , however, is seen to vary considerably across subrounds. The coefficients of old persons for both Poisson and Negative Binomial models are found to be non-significant for Subround 4 *and* those for children are found to be non-significant for Subround 1. Similarly, education seems to have no significant effect in Subrounds 1 and 4. [4]

Finally, we note that the estimated standard errors of some covariates in the model turn out to be very high. Many coefficients, in fact, turn out to be non-significant. For example, the coefficients of WEST and FEMALE are non-significant for all the subrounds for both Poisson and Negative Binomial model. Given the wide regional variations in Maharashtra, the non-significance of the coefficient WEST for all the four subrounds in both Poisson and Negative Binomial models is somewhat surprising. Also, it is generally observed that the female members of the households enter and exit from the labour market more frequently (Bardhan, 1984). However, such movements generally take place *across* agricultural seasons. Our results imply that *within* the 'peak' season or within the 'lean' season, the extent of movements of the females may not be significantly higher than that of the males. Still, it will be interesting to examine whether this same phenomenon is observed for other states of India.

3.3.4 Goodness of Fit

We now check whether the specification of the above models indeed leads to good results. We also draw comparisons between the Poisson and Negative Binomial models in terms of goodness of fit to the data. It is customary to use the chi-square goodness of fit test for this purpose, which compares the observed and the expected probabilities of the cells in the model. However, here the possible values taken by the variable is only seven, whereas there are

[4]Note that the observed t-ratios for one sided and two sided tests at 5% level of significance should be greater than 6.314 and 12.706 respectively.

nine parameters in the model. Therefore, the chi-square test could not be conducted for the combined sample. One can of course increase the number of cells by separately looking across the various covariate groups. However, the total number of possible covariate combinations is around one hundred. Since the sample size in each subround is not more than seven thousand, there will be many cells with only a few observations and this will make the chi-square test invalid. It may also be noted that for large samples, the standard chi-square goodness of fit test very often leads to rejection of the proposed models, because with increase in sample size the chi-square test demands much closer fit than is possible in most real life problems. We, therefore, measure goodness of fit by other criteria.

We consider an alternative approach followed by Cooil(1991) in which the distributional assumption is checked directly by examining whether the models are able to generate the right predictive frequencies for each covariate group. Let p_j represent the observed proportion of persons in any subround who have made j jumps in the reference period. We define the corresponding model estimates as the average unconditional predictive probabilities, i.e.,

$$\widehat{p_j} = \bar{f}(j) = \sum_{i=1}^{n} w_i \hat{f}_i(j) \tag{3.3.3}$$

Here $\hat{f}_i(j)$ is the unconditional probability implied by the model for the i-*th individual to jump j times in the reference period.* These probabilities are averaged across all individuals, using the corresponding multipliers as weights. The graphs of observed and the estimated probabilities are presented subroundwise in Figure 3.3.1.

From Figure 3.3.1 it is clear that the Negative Binomial model provides a better fit to the data for all four subrounds. We shall, however, compare the performance of the proposed models by some summary measures of goodness of fit. Following Cooil(1991), the basic measures of predictive fit are taken as :

(a) the proportional prediction error (PPE_0) which gives the relative error in fit corresponding to the probability of zero jumps

$$\mathrm{PPE}_0 = \frac{|\, p_0 - \hat{p}_0 \,|}{\hat{p}_0} \tag{3.3.4}$$

(b) the total absolute prediction error (TAPE) made in estimating the cell

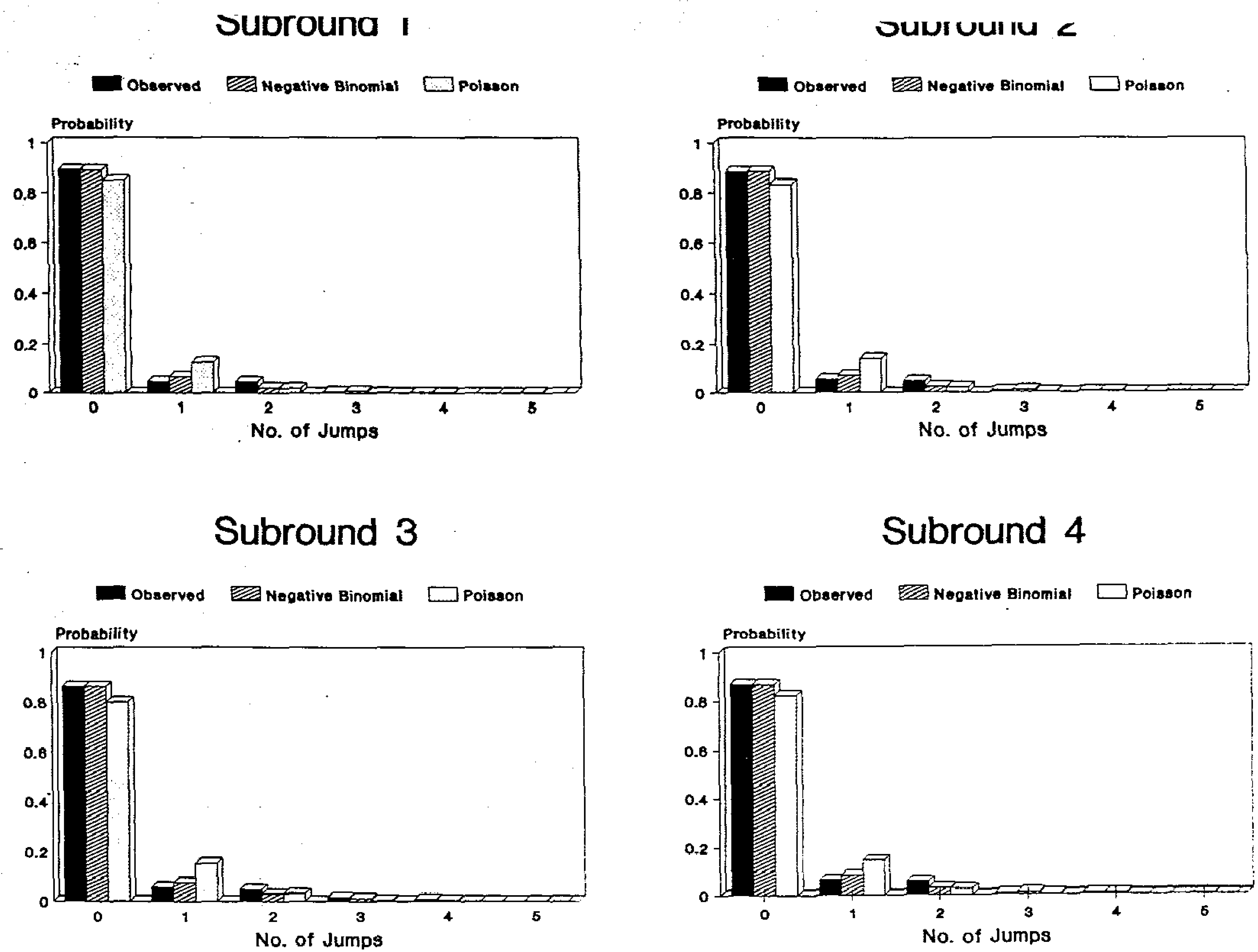

Figure 3.3.1 : Observed and Estimated Proportions of Jumps for Poisson and Negative Binomial Model accross Subrounds: Rural Maharashtra, NSSO 38-th Round Combined Sample.

probabilities, i.e.,

$$\text{TAPE} = \sum_j \mid p_j - \hat{p}_j \mid \tag{3.3.5}$$

and also

(c) the maximum distance between the observed and the predicted cumulative relative frequencies, i.e.,

$$D = \max_{\{y\}} \mid F_n(y) - F(y) \mid \tag{3.3.6}$$

where $F_n(y)$ is the empirical distribution function (EDF) and $F(y)$ is the proposed distribution function.

Note that for all the four subrounds, the expected probabilities do not sum to one for the Negative Binomial model. This is because the Negative Binomial distribution has a long tail. It is the long tail of this distribution which tackles the problem of overdispersion more efficiently than a Poisson model. Therefore, the expected probabilities for the cells do not sum to one as there exists some positive probability for more than six jumps! However, these probabilities are small for all subrounds so that it may be considered as a fair approximation. Generally, the sum of the expected probabilities are made unity by adjusting the remaining probabilities proportionately to all the cells. We have not done that since this will not reveal the extent of prediction error outside the relevant sample range committed by the model. Instead we add this error in our measurement of TAPE. For Poisson model, however, such errors do not occur as the p_j very quickly converges to zero as j increases. We present the calculated values of PPE_0 and TAPE in Table 3.3.3.

While the values of TAPE for the Poisson model lie between 0.15 and 0.20, the corresponding range for the Negative Binomial model is between 0.045 and 0.055 only. The Poisson model grossly underestimates the probability of no jump and grossly overestimates the probability of one jump. Compared to Poisson model, the performance of Negative Binomial model is far more superior. The error in estimation in the probability of no jump is very little for Negative Binomial model for all subrounds. The tail probabilities are also better estimated by Negative Binomial model. However, for all subrounds, the probability of one jump is slightly overestimated and the corresponding probability for two jumps is slightly underestimated by the Negative Binomial model. The main contribution to TAPE for Negative

Table 3.3.3: Obtained Values of Some Goodness of Fit Measures for the Poisson and the Negative Binomial Models : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Model	Measure	Subround 1	Subround 2	Subround 3	Subround 4
(1)	(2)	(3)	(4)	(5)	(6)
	PPE_0	0.0489	0.0557	0.0700	0.0515
Poisson	TAPE	0.1556	0.1676	0.1909	0.1625
	D	0.0437	0.0492	0.0605	0.0447
	PPE_0	0.0020	−0.0011	−0.0006	0.0002
Negative	TAPE	0.0524	0.0451	0.0469	0.0541
Binomial	D	0.0182	0.0171	0.0185	0.0194

Binomial model comes from the discrepancies in the observed and predicted probabilities for these two cells only.

3.4 Conclusion

This chapter examined the extent of employment fluctuation for the Indian rural sector as reflected by the NSSO 38-th round household level data on employment and unemployment for rural Maharashtra. We measured day-to-day employment fluctuation by the number of jumps observed for different individuals across several employment states within a moving reference period of seven days. The statistical models used for this purpose were essentially generalisations of Poisson and Negative Binomial distributions. Empirical investigations carried out on the basis of these models confirmed the existence of substantial heterogeneity in the pattern of employment fluctuation across different types of households.

Broadly speaking, we found that the persons from wage dependent households in rural India are substantially more susceptible to employment fluctuation than the rest of the population. For the highly educated persons, on the other hand, there are fewer state changes compared to others. Age and the number of jumps are found to be very highly related, but we have found no significant relationship between number of jumps and other covariates (e.g., region, social group and sex). It is, however, not unlikely that some degree of interaction effects between covariates exists which have not been explored. It would be worthwhile to examine the presence of such interaction effects.

We now propose a few generalisations. Although the number of jumps effectively summarises the extent of employment fluctuation, some amount of information is lost because of this summarisation. A detailed study which distinguishes transitions across several employment states will be more helpful to trace the source of this fluctuation.

More importantly, in this chapter we observed each individual for seven days only. Since the reference period of data collection corresponding to an individual is so short, the number of jumps observed for an individual may depend on his initial state. Convergence to stationary steady state employment may take some more time. To remove this dependence on the initial state, each individual will have to be observed for a large number of days. With the help of the analytical tools described in this chapter, such an analysis should be possible which may throw up some more light on the problem of employment fluctuation.

Employment fluctuation may be measured in a number of possible ways, one aspect of such a measure being *the extent of movement* and another being *the extent of predictability.* The number of jumps from one employment state to another is basically concerned with the extent of movement. In Chapters 4 and 5 of this dissertation, we shall be concerned with the other aspect, viz., predictability. Of these two, Chapter 4 specifies a model where employment of an individual is assumed to follow independent Markov chains. Chapter 4 finds out the conditional distribution of the 'current' aggregate employment vector given the aggregates of the immediately preceding period. Chapter 5 then interprets the trace of the dispersion matrix of this conditional distribution as a possible measure of employment fluctuation and applies it to the NSSO 38-*th* round data on employment and unemployment for rural Maharashtra. Since the proposed measure is based on the dispersion matrix, it reveals the extent of predictability of the employment states of individuals. Thus Chapters 2, 3, 4 and 5, taken together will provide a more complete picture of employment fluctuation in rural India.

Chapter 4

Determination of Aggregate Employment Fluctuation

4.1 Introduction

In Chapters 2 and 3 we analysed individuals' employment fluctuation measured in terms of the observed number of jumps across different employment states during the reference period. Although the number of jumps effectively summarises employment fluctuation, some amount of information is lost as a result of this summarisation. In this chapter we shall examine the phenomenon of employment fluctuation with the help of a conventional Markov chain model.

To study the phenomenon of employment fluctuation in a Markov chain framework, one should ideally use a continuous cross-sectional (panel) data on transition of employment states of n individuals for T periods. However, often in real life, instead of such panel data, all one gets are the data aggregated over individuals. The models which are used to analyse such data are broadly of two types : (a) models which take extensive help of economic theory (Andrews, 1987) *and* (b) statistical time series models (Funke, 1992). Of these, models under the first category often specify a simultaneous equations framework. The practice of building up large structural models sequentially, equation by equation, has been criticised by Sims(1980) because these models often impose arbitrary exclusion restrictions for the purpose of identification. As an alternative, the use of atheoretical time series models has been prescribed by Sims.

The nature of specification of a statistical time series model depends on the level of aggregation of the available data. If the data are aggregated over individuals, a multivariate time series model may be specified. Of special interest are the vector autoregressive moving average (VARMA) models as these are sufficiently general in the sense that they provide a fairly good approximation to the reduced form of any structural system of simultaneous equations (Zellner and Palm, 1974; Zellner, 1979). Under fairly general conditions, these models can be written as *either* vector autoregressive (VAR) models *or* vector moving average (VMA) models. The univariate AR(p), MA(q), or ARMA(p,q) models are special cases of VARMA models.

When the data are available at a fairly disaggregated level, e.g., when we have the data on aggregate employment level of counties or districts for T periods separately, spatial-time series models may be specified (Bronars and Jansen, 1987; Dunn, 1987). The scope of spatial-time series models, however, is somewhat limited because such models require the space to be divided into grids equal in area or size and the aggregate measures of the concerned variable are to be calculated for each of these grids .

This chapter examines the aggregation implication of the VARMA models. It is obvious that the period to period values of aggregate employment form a time series. If only this aggregate information is known, one uses an AR(p), MA(q) or ARMA(p,q) process (Funke, 1992). These models are ad hoc in the sense that they do not have any micro-theoretic foundation. In contrast, the model proposed in this chapter starts from an individual specific assumption, viz., that the period to period changes in the state of employment of individuals follow independent Markov chains.

Section 4.2 of this chapter presents a general analytical framework. On any day of the reference week, an individual may either be employed, unemployed or be out of the labour force. Thus the total number of possible states in this problem is only three. The analytical framework developed, however, assumes the number of states over which movements of individuals may occur to be K. This type of structure occurs frequently in economics and in other social sciences (Bartholomew, 1982). The advantage of specifying this general framework is that it can be applied to similar type of problems in other research areas.[1]

[1]See Frydman (1984) and Sampson (1990), which considers mobility across different types

Section 4.3 of this chapter finds out the conditional distribution of the state aggregates over individuals, given the aggregates of the corresponding 'current' states. In Section 4.4, we consider the special case of $K = 2$ in which the state aggregates over individuals generate a *univariate* time series model the solution of which is very simple and has an easy interpretation. Section 4.5 considers an empirical application of the proposed framework to a study of the nature and the extent of fluctuation of the aggregate employment series for rural Maharashtra. Finally, Section 4.6 summarises the main findings and makes some concluding observations.

4.2 The General Analytical Framework

Let X_{it}^k be a random variable which denotes whether the i-th individual was in the k-th state on the t-th day. We define,

$$
\begin{aligned}
X_{it}^k &= 1 \qquad \text{if the } i\text{-th individual was in the } k\text{-th state} \\
& \qquad\quad\ \text{on the } t\text{-th day} \\
&= 0 \qquad \text{otherwise}
\end{aligned} \tag{4.2.1}
$$

Since there are K states, we define

$$\tilde{X}_{it}' = (X_{it}^1, X_{it}^2, \ldots, X_{it}^K) \tag{4.2.2}$$

where $\tilde{X}_{it}$ is the $K \times 1$ vector giving the position of the i-th individual on the t-th day. One and only one component of $\tilde{X}_{it}$ is one, the rest being zero, so that

$$\sum_{k=1}^{K} X_{it}^k = 1, \qquad \text{each } X_{it}^k = 0 \text{ or } 1 \tag{4.2.3}$$

Let $\mathbf{X}_i$ denote observations corresponding to the i-th individual for T successive days, i.e.,

$$\mathbf{X}_i = (\tilde{X}_{i1}, \tilde{X}_{i2}, \ldots, \tilde{X}_{iT}) \qquad i = 1, 2, \ldots, n \tag{4.2.4}$$

which is a $K \times T$ matrix. It shows the daily state positions of the individual i throughout the reference period. We assume that for fixed i, $\tilde{X}_{it}$ follows

of jobs and Conlisk (1989), Conlisk (1990), Coondoo and Dutta (1992) and Dardanoni (1993), which study individuals' movements across income classes.

a Markov chain, but across i (which means, across individuals), $\widetilde{X}_{it}$'s are independent and identically distributed (*iid*).

We also assume that for fixed i, $\widetilde{X}_{it}$ is stationary in the strict sense, i.e.,

$$F(\widetilde{X}_{i1}, \widetilde{X}_{i2}, \ldots, \widetilde{X}_{is}) = F(\widetilde{X}_{i(t_0+1)}, \widetilde{X}_{i(t_0+2)}, \ldots, \widetilde{X}_{i(t_0+s)}) \qquad (4.2.5)$$

where $F(\cdot)$ is the cumulative distribution function (CDF) of $(\widetilde{X}_{i1}, \widetilde{X}_{i2}, \ldots, \widetilde{X}_{is})$.

We assume that $\mathbf{E}(\widetilde{X}_{it})$ and $\mathbf{D}(\widetilde{X}_{it})$ exist where $\mathbf{E}$ and $\mathbf{D}$ denote the expectation operator and the dispersion matrix respectively. Because of strict stationarity, $\mathbf{E}(\widetilde{X}_{it})$ and $\mathbf{D}(\widetilde{X}_{it})$ will be independent of t. Also since $\widetilde{X}_{it}$'s are *iid* across i, $\mathbf{E}(\widetilde{X}_{it})$ and $\mathbf{D}(\widetilde{X}_{it})$ will be independent of i as well.

Let us denote

$$\mathbf{Y} = \sum_{i=1}^{n} \mathbf{X}_i \qquad (4.2.6)$$

and

$$\mathbf{Z} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \qquad (4.2.7)$$

Then $\mathbf{Y}$ and $\mathbf{Z}$ are two $K \times T$ matrices which reflect aggregates and proportions of cell positions respectively for each state across the whole reference period. Let us define

$$\mathbf{Y} = (\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_T) \qquad (4.2.8)$$

and

$$\mathbf{Z} = (\widetilde{Z}_1, \widetilde{Z}_2, \ldots, \widetilde{Z}_T) \qquad (4.2.9)$$

Thus, $(\widetilde{Y_1}, \widetilde{Y_2}, \ldots, \widetilde{Y_T})$ or $(\widetilde{Z_1}, \widetilde{Z_2}, \ldots, \widetilde{Z_T})$ gives us a multivariate time series. Here $\widetilde{Y}_t$ and $\widetilde{Z}_t$ give us aggregate and proportionate state position on day t respectively. We denote the k-th component of $\widetilde{Y}_t$ and $\widetilde{Z}_t$ by Y_t^k and Z_t^k respectively. Here, Y_t^k measures the aggregate number of individuals in the k-th state on the t-th day. A similar interpretation can be given to Z_t^k.

Let the initial probability vector be given by

$$\widetilde{\pi}^{0'} = (\pi_1^0, \pi_2^0, \ldots, \pi_K^0) \qquad (4.2.10)$$

where π_k denotes the initial probability of being in the k-th state. Further, let the transition probability matrix be

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ \cdots & \cdots & \cdots & \cdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{pmatrix} \qquad (4.2.11)$$

where p_{kj}'s are the transition probabilities, i.e.,

$$p_{kj} = P\,[X^j_{i(t+1)} = 1 \mid X^k_{it} = 1] \tag{4.2.12}$$

We assume that the transition probabilities are constant across time. For convenience, we shall denote the k-th row of the transition probability matrix $\mathbf{P}$ by

$$P^{(k)'} = (p_{k1}, p_{k2}, \ldots, p_{kK}) \qquad k = 1, 2, \ldots, K \tag{4.2.13}$$

Thus the k-th row of the transition probability matrix corresponds to a multinomial probability distribution with the probability vector $P^{(k)'}$.

Let the steady state distribution be given by

$$\widetilde{\pi}^{*'} = (\pi_1^*, \pi_2^*, \ldots, \pi_K^*) \tag{4.2.14}$$

This steady state distribution is obtained by solving the equation

$$\widetilde{\pi}^{*'} = \widetilde{\pi}^{*'}\mathbf{P} \tag{4.2.15}$$

Note that one equation in the above system of linear equations will be redundant and should be replaced by the condition

$$\sum_{k=1}^{K} \pi_k^* = 1 \tag{4.2.16}$$

Let us denote

$$\mathbf{V} = \mathrm{Diag}(\pi_1^*, \pi_2^*, \ldots, \pi_K^*) - \widetilde{\pi}^*\widetilde{\pi}^{*'} \tag{4.2.17}$$

Note that each row and each column of $\mathbf{V}$ sums to zero. Therefore, the rank of $\mathbf{V}$ is at most $(K-1)$.

We now have the following proposition.

Proposition 4.2.1 $\widetilde{Z}_t \xrightarrow{\mathcal{L}} \widetilde{Z}^*$ *where* $\widetilde{Z}^* \sim N(\widetilde{\pi}^*, \frac{1}{n}\mathbf{V})$ *and* $\xrightarrow{\mathcal{L}}$ *denotes convergence in distribution.*

Proof : Since $\widetilde{X}_{it}$ is stationary for fixed i,

$$\widetilde{X}_{it} \xrightarrow{\mathcal{L}} \widetilde{X}_i^* \qquad \text{as } t \to \infty \tag{4.2.18}$$

where $X_i^* \sim \mathrm{Multinomial}(1, \widetilde{\pi}^*)$.

Since $(\widetilde{X}_{it},\ i = 1, 2, \ldots, n)$ are independent,

$$\widetilde{Z}_t = \frac{1}{n}\sum_{1=1}^{n} \widetilde{X}_{it} \xrightarrow{\mathcal{L}} \frac{1}{n}\sum_{i=1}^{n} \widetilde{X}_i^* \tag{4.2.19}$$

Clearly,

$$\sum_{i=1}^{n} \widetilde{X}_i^* \sim \text{Multinomial}(n, \widetilde{\pi}^*) \tag{4.2.20}$$

As n is very large, by central limit theorem (CLT),

$$\frac{1}{n}\sum_{i=1}^{n} \widetilde{X}_i^* \xrightarrow{\mathcal{L}} \widetilde{Z}^* \tag{4.2.21}$$

where $Z^* \sim N(\widetilde{\pi}^*, \frac{1}{n}\mathbf{V})$. Therefore, $\widetilde{Z}_t \xrightarrow{\mathcal{L}} \widetilde{Z}^*$ where $\widetilde{Z}^* \sim N(\widetilde{\pi}^*, \frac{1}{n}\mathbf{V})$. ∎

Corollary 4.2.1 *For $K = 2$, $Z_t \xrightarrow{\mathcal{L}} Z^*$ where $Z^* \sim N[\pi^*, \frac{1}{n}\pi^*(1 - \pi^*)]$*

Note that in Corollary 4.2.1 both the mean and the variance of Z^* are functions of the steady state probability. Also, $\text{Var}(Z^*)$ is maximum if $\pi^* = 1/2$. This is the case when the univariate time series of any of the labour states of employment, unemployment or out of labour force fluctuates most. At the other end, $\text{Var}(Z^*) = 0$ if $\pi^* = 0$ *or* $\pi^* = 1$. For these cases, Z^* is degenerate.

4.3 Derivation of the Conditional Distribution

The marginal distribution of $\widetilde{Z_t}$ may not yield good predictions as the variance of $\widetilde{Z_t}$ may be quite large. However, use of the conditional distribution of $\widetilde{Z_t}$ given $\widetilde{Z_{t-1}}$ will vastly improve the efficiency of prediction. Let

$$\widetilde{X}_{i(t+1)} = \sum_{k=1}^{K} \widetilde{X}_{it}^k V_{it}^k \tag{4.3.1}$$

where $V_{it}^k \sim \text{Multinomial}(1, P^{(k)})$. The intuitive justification of the above equation is as follows : the probability distribution of $\widetilde{X}_{i(t+1)}$ depends on $\widetilde{X}_{it}$. Since, $\widetilde{X}_{it}$ is a $K \times 1$ vector and one and only one component of $\widetilde{X}_{it}$ is unity (the rest being zero), $\widetilde{X}_{i(t+1)} \sim \text{Multinomial}(1, P^{(k)})$ if and only if $X_{it}^k = 1$.

Note that, V_{it}^k, $(k = 1, 2, \ldots, K)$ are not independent. However, if the value of $\widetilde{X}_{it}$ is known, then $V_{it}^k, (k = 1, 2, \ldots, K)$ are independent. Aggregating across i, we obtain

$$\widetilde{Y}_{t+1} = \sum_{i=1}^{n} X_{i(t+1)} = \sum_{i=1}^{n} \sum_{k=1}^{K} X_{it}^k V_{it}^k \tag{4.3.2}$$

Interchanging summations, we write the above equation as

$$\widetilde{Y}_{t+1} = \sum_{k=1}^{K} \sum_{i=1}^{n} X_{it}^k V_{it}^k = \sum_{k=1}^{K} V_k \tag{4.3.3}$$

where

$$V_k = \sum_{i=1}^{n} X_{it}^k V_{it}^k \tag{4.3.4}$$

Clearly,

$$V_k \mid \widetilde{Y}_t \sim \text{Multinomial}(Y_t^k, P^{(k)}) \tag{4.3.5}$$

Note that $V_1, V_2, \ldots, V_K$ are not independent. However, it can be easily proved that if the vector $\widetilde{Y}_t$ is known, then $V_1, V_2, \ldots, V_K$ are independent.

Now if n is very large so that $(Y_t^k, k = 1, 2, \ldots, K)$ are all very large, we can approximate each multinomial by a multivariate normal probability distribution. Note that since $\sum_{j=1}^{K} p_{kj} = 1$ for all $k = 1, 2, \ldots, K$, the central limit approximation will yield a $(K-1)$ dimensional normal random variable.

Let,

$$\mathbf{U}_k = \text{Diag}(p_{k1}, p_{k2}, \ldots, p_{kK}) - P^{(k)} \cdot P^{(k)'} \tag{4.3.6}$$

Then,

$$\frac{1}{n} V_k \sim N(Z_t^k P^{(k)}, \frac{1}{n} Z_t^k \mathbf{U}_k) \qquad \text{as } n \to \infty \tag{4.3.7}$$

Since given $\widetilde{Z}^t$, the random variables $V_1, V_2, \ldots, V_K$ are independent, their central limit approximations will also behave like independent random variables. Therefore,

$$\widetilde{Z}_{t+1} \mid \widetilde{Z}_t \sim N(\sum_{k=1}^{K} Z_t^k P^{(k)}, \frac{1}{n} \sum_{k=1}^{K} Z_t^k \mathbf{U}_k) \tag{4.3.8}$$

Thus (4.3.8) follows a nonstationary VAR process with drift, the nonstationarity entering through the dispersion matrix of the disturbance vector.

The result presented above may be interpreted as follows : when $\widetilde{Z}_t$ is not known, we can only use the fact that the distribution of $\widetilde{Z}_{t+1}$ will be a $(K-1)$

dimensional multivariate normal. Both the mean vector and the dispersion matrix of that distribution will be functions of the steady state probabilities. However, when $\widetilde{Z}_t$ is known, not only do we have more information regarding the average, we also have more information regarding the *fluctuation* of the process. If there exists wide differences in the extent of fluctuation across rows of the transition matrix $\mathbf{P}$, naturally that information will manifest in the conditional distribution. For example, let us consider two rows of $\mathbf{P}$ (say, the first and the second), one being $(1, 0, 0, \ldots, 0)$ and the other being $(\frac{1}{K}, \frac{1}{K}, \cdots, \frac{1}{K})$. If it is *known* that currently N_1 observations are in the first cell and N_2 observations are in the second, then not only do we know the cell-averages, but we also know that the above N_1 persons will remain in the first cell in the next period *with certainty*. This additional information will reduce the variance of the conditional distribution.

4.4 The Case K=2 : A Special Case

This section discusses the special case of K=2. Although the results derived in Sections 4.2 and 4.3 are very general, the case $K = 2$ should be studied separately for several reasons. When K=2, the state aggregates over individuals generate a *univariate* time series. The fluctuation of this univariate series can be very easily explained by the transition probabilities. The resulting solution is very simple and can be easily interpreted. In fact, the state aggregates follow a nonstationary AR(1) process with drift. Both the mean and the variance of this process are functions of the transition probabilities of the corresponding Markov chain. Also, since the parameters of this process are related to the transition probabilities, several important types of fluctuation of the state aggregates can be linked up with the transition probabilities more intensively. In Section 4.5, we shall consider an empirical illustration of this special case. We shall study the extent and the nature of fluctuation of the aggregate employment series for rural Maharashtra.

For the special case of K=2, we simplify the notation a bit. The notation adopted in this section supersede the old notation and are maintained throughout the rest of this chapter. Also for easy reference, we shall call the states as *employment* (E) and *non-employment* (N).

Let X_{it} be a random variable which denotes the employment position of the i-th individual on the t-th period ($i = 1, 2, \ldots, n$; $t = 1, 2, \ldots, T$). For convenience, we shall call each such period a *day*. We define,

$$\begin{aligned} X_{it} &= 1 \quad \text{if the } i\text{-th individual is employed on } t\text{-th day} \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{4.4.1}$$

Let us denote

$$\widetilde{Y} = \sum_{i=1}^{n} \widetilde{X}_i \tag{4.4.2}$$

and

$$\widetilde{Z} = \frac{1}{n} \sum_{i=1}^{n} \widetilde{X}_i \tag{4.4.3}$$

Here, $\widetilde{Y}$ and $\widetilde{Z}$ represent the vector of aggregate (i.e., the total number of employed persons) and the proportion of employed persons on each day of the reference period. Let us denote $Y_t = \sum_{i=1}^{n} X_{it}$ and $Z_t = \frac{1}{n} \sum X_{it}$. Here, $(Y_1, Y_2, \ldots, Y_T)$ *or* $(Z_1, Z_2, \ldots, Z_T)$ gives us a time series; Y_t gives us the aggregate employment on day t and Z_t gives us the proportion of people employed on day t respectively.[2]

Let the initial probability vector be

$$\widetilde{\pi}_0{}' = (\pi_0, 1 - \pi_0) \tag{4.4.4}$$

where π_0 is the initial probability of employment.

The transition matrix is given by

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \tag{4.4.5}$$

where $p = P[X_{t+1} = 1 \mid X_t = 1]$ and $q = P[X_{t+1} = 0 \mid X_t = 0]$.

Also, let the steady-state distribution be given by

$$\widetilde{\pi}^{*\prime} = (\pi^*, 1 - \pi^*) \tag{4.4.6}$$

Note that π^* is obtained by solving the equation

$$\pi^* p + (1 - \pi^*)(1 - q) = \pi^* \tag{4.4.7}$$

[2]Strictly speaking, since the reference period for data collection may be different across different individuals $\sum_{i=1}^{n} X_{it}$ or $\frac{1}{n} \sum X_{it}$ may not reflect the actual employment position on the t-th day literally, but under strict stationarity its distribution will be the same as the above.

Let π_t be the probability that the i-th individual will be employed on the t-th day. Then it can be easily proved that

$$\pi_t = P[X_{it} = 1] = \pi^* + (p + q - 1)^t (\pi_0 - \pi^*) \tag{4.4.8}$$

Corollary 4.4.1 *If Y_t and $(n - Y_t)$ are both very large, then,*

$$Z_{t+1} \mid Z_t \sim N[Z_t p + (1 - Z_t)(1 - q), \frac{Z_t p(1-p)) + (1 - Z_t)q(1-q)}{n}]$$

Note that we can write,

$$Z_{t+1} = (1 - q) + (p + q - 1)Z_t + \epsilon_{t+1} \tag{4.4.9}$$

where ϵ_{t+1} is a random variable with

$$\mathbf{E}(\epsilon_{t+1} \mid Z_t) = 0 \tag{4.4.10}$$

and

$$\text{Var}\,(\epsilon_{t+1} \mid Z_t) = \frac{Z_t p(1-p) + (1 - Z_t).q(1-q)}{n} \tag{4.4.11}$$

Therefore Z_t follows a nonstationary AR(1) process with a constant drift, the nonstationarity entering through variance of the disturbance term. Note that

$$\mathbf{E}(\epsilon_{t+1}) = \mathbf{E}_{Z_t}.(\mathbf{E}(\epsilon_{t+1} \mid Z_t)) = 0 \tag{4.4.12}$$

and

$$\mathbf{E}(Z_{t+1}) = p \cdot \mathbf{E}(Z_t) + (1 - q) \cdot (1 - \mathbf{E}(Z_t)) \tag{4.4.13}$$

Also,

$$\lim_{t\to\infty} \mathbf{E}(Z_{t+1}) = \lim_{t\to\infty} \mathbf{E}(Z_t) = \pi^* \tag{4.4.14}$$

This means, as $t \to \infty$, (4.4.13) becomes the steady-state equation.

Subtracting (4.4.13) from (4.4.9), we obtain

$$(Z_{t+1} - \mathbf{E}(Z_{t+1})) = (p + q - 1)(Z_t - \mathbf{E}(Z_t)) + \epsilon_{t+1} \tag{4.4.15}$$

Since $\mid p + q - 1 \mid < 1$, if $Z_t - \pi_t \geq 0$, there will be tendencies to push it down to the steady state level, the extent of adjustment depending upon the value of $\mid p + q - 1 \mid$. If $\mid p + q - 1 \mid$ is close to zero, the adjustment will be quick, if $\mid p + q - 1 \mid$ is close to one, adjustment will be slow. Note that

$(p+q-1)$ is an eigenvalue of the transition matrix $\mathbf{P}$. Since here $\mathbf{P}$ is a 2×2 matrix there can be at most two eigenvalues of $\mathbf{P}$. The other eigenvalue of $\mathbf{P}$ is unity. If $\mid p+q-1 \mid < 1$, it will be the *second highest eigenvalue* of $\mathbf{P}$. The second highest eigenvalue is intimately related to the mobility inherent in a transition matrix. Measures of mobility which are functions of the second highest eigenvalue reflect how quickly convergence takes place to the steady state (Shorrocks, 1978).

The mechanism of adjustment not only depends upon the value of $\mid p+q-1 \mid$, but also upon the difference between $p(1-p)$ and $q(1-q)$. Here $p(1-p)$ is the expression of variance of Bernoulli (p) and similarly $q(1-q)$ is the variance of Bernoulli $(1-q)$. If this difference is zero, both the states E and N are similar so far as fluctuation is concerned and hence variance of aggregate employment will be free from Z_t. However, if this gap is large, and we know the number of persons in states E and N on a given day, we can utilise this information for better prediction. We thus have the following proposition :

Proposition 4.4.1 *The conditional distribution of Z_{t+1} given Z_t is free from Z_t if and only if the transition matrix is of the form :*

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ p & 1-p \end{pmatrix} \tag{4.4.16}$$

Proof : Clearly, the conditional distribution will be free from Z_t if and only if

$$p+q=1 \tag{4.4.17}$$

and

$$p(1-p)=q(1-q) \tag{4.4.18}$$

As (4.4.17) implies (4.4.18), (4.4.18) becomes redundant.

Thus, the conditional distribution of $Z_{t+1} \mid Z_t$ will be free of Z_t if and only if the transition matrix $\mathbf{P}$ is of the form (4.4.16). ■

The intuition of the above result is as follows : if two states are identical with respect to transition, then the additional information of knowing Z_t will not be of any extra help.

Corollary 4.4.2 *Z_{t+1} follows a stationary $AR(1)$ process if $p(1-p) = q(1-q)$.*

Proof : Follows directly from (4.3.8). ∎

The intuitive reasoning behind this result is that since the variance within each row of P is the same, aggregate variance will be free of the number of persons in each state.

We now discuss a few important special cases.

Case 1 : $p + q \approx 2$

In this case, both p and q are close to 1. This suggests that the market is comparatively less shock prone because variance across days is small. However, if a shock displaces employment rate from the steady state solution, the market adjusts to its previous level slowly but steadily.

Case 2 : $p + q > 1, p \approx q \approx 1/2$

In this case the conditional variance is high. The market is more shock prone. However, adjustment takes place quickly.

Case 3 : $p + q > 1, p \approx 1, q \approx 0$

Here the market can adjust very fast. Also, conditional variance will be lower compared to other cases, hence convergence is also likely to be steady.

Case 4 : $p + q > 1, p \approx 0, q \approx 1$

Similar as Case 3.

Case 5 : $p + q < 1$

Since both p and q are small, high unemployment on one day is associated with low employment on the next. This case is very unlikely to occur in reality.

4.5 An Empirical Illustration

In this section, we illustrate the theoretical results derived in the previous sections with an example. Since this is an illustration, we shall work with the two states proposed in Section 4.4, viz., employment (E) and non-employment (N). This means, we shall not distinguish between the states unemployment (U) and out of labour force (O).[3]

Since the transition probabilities may not remain constant across subrounds, the initial and the transition probabilities are separately estimated across subrounds.[4] The results obtained are presented in Table 4.5.

Table 4.5.1: Estimates of the Initial, Transition and Steady State Probabilities across Subrounds : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Estimate of (1)	Subround 1 Jan-Mar (2)	Subround 2 Apr-Jun (3)	Subround 3 Jul-Sep (4)	Subround 4 Oct-Dec (5)
Initial Probability (π_0)	0.4599 (0.0026)*	0.4191 (0.0031)	0.4800 (0.0016)	0.4743 (0.0002)
Transition Probability from E to E (p)	0.9683 (0.0018)	0.9628 (0.0016)	0.9586 (0.0054)	0.9659 (0.0026)
Transition Probability from N to N (q)	0.9720 (0.0014)	0.9698 (0.0000)	0.9594 (0.0042)	0.9624 (0.0010)
Steady State Probability (π^*)	0.4690 (0.0022)	0.4485 (0.0105)	0.4954 (0.0067)	0.5246 (0.0125)
Log-likelihood	−10828.62	−10894.55	−10098.84	−11573.98
Sample Size	6162	6203	5032	6077

* The bracketed terms are estimates of standard errors.

For all the four subrounds, we find that p and q are very close. The maximum difference between $p(1-p)$ and $q(1-q)$ occurs for Subround 2, the value being 0.0065. On the other hand, for Subround 3 it is closest to zero, the value being 0.0007. Since the values of the difference are so close in all the cases, we conclude that a stationary AR(1) process with a drift approximates the 'true' distribution well.

For each day in each subround, we now compute the estimated value of

[3]The difference between U and O is, however, very thin (Tano, 1991).

[4]For a discussion on the method of estimation, see Appendix A.

the 'current' employment rate given the information of the preceding period, i.e.,

$$\widehat{Z_{t+1}} = (1-\widehat{q}) + (\widehat{p}+\widehat{q}-1)Z_t \tag{4.5.1}$$

We also compute the estimated standard error $\widehat{\sigma_{t+1}}$, where

$$\widehat{\sigma_{t+1}}^2 = \{\frac{1}{n}[\widehat{p}(1-\widehat{p})Z_t + \widehat{q}(1-\widehat{q})(1-Z_t)]\} \tag{4.5.2}$$

Note that if $\widehat{p}(1-\widehat{p})$ and $\widehat{q}(1-\widehat{q})$ are very close, then

$$\widehat{\sigma_{t+1}} \approx [\widehat{p}(1-\widehat{p})]^{\frac{1}{2}} \approx [\widehat{q}(1-\widehat{q})]^{\frac{1}{2}} \tag{4.5.3}$$

i.e., $\widehat{\sigma_{t+1}}$ is free from Z_t. The above condition holds in this case. For the four subrounds the estimated standard errors are 0.0022, 0.0023, 0.0028 and 0.0024 respectively.

The observed and the estimated values are presented in Figure 4.5.1. Figure 4.5.1 reveals that the observed and the estimated values are quite close. The maximum absolute deviation is less than 0.01. For most of the observations, the observed values fall within the 95% confidence interval of the estimated values.

4.6 Conclusion

This chapter derived the aggregation implication of the VARMA models from individual specific assumptions. Under the assumption that the state vectors of individuals follow stationary and *iid* Markov chains, the conditional distribution of current cell aggregates given the cell aggregates of the preceding period has been derived. The conditional distribution was shown to be a nonstationary VAR process. We also obtained the conditions under which this process would be stationary. Empirical examinations with NSSO 38-th round data on employment revealed that a stationary AR(1) process provides a good approximation for the employment series.

Note that throughout this chapter we worked with *iid* Markov chains. So far as the empirical illustration is concerned, this is a serious limitation – the transition probabilities may not be constant across individuals. It is easy to generalise the model proposed here and such a generalisation may be attempted from two directions : one can consider mover-stayer models

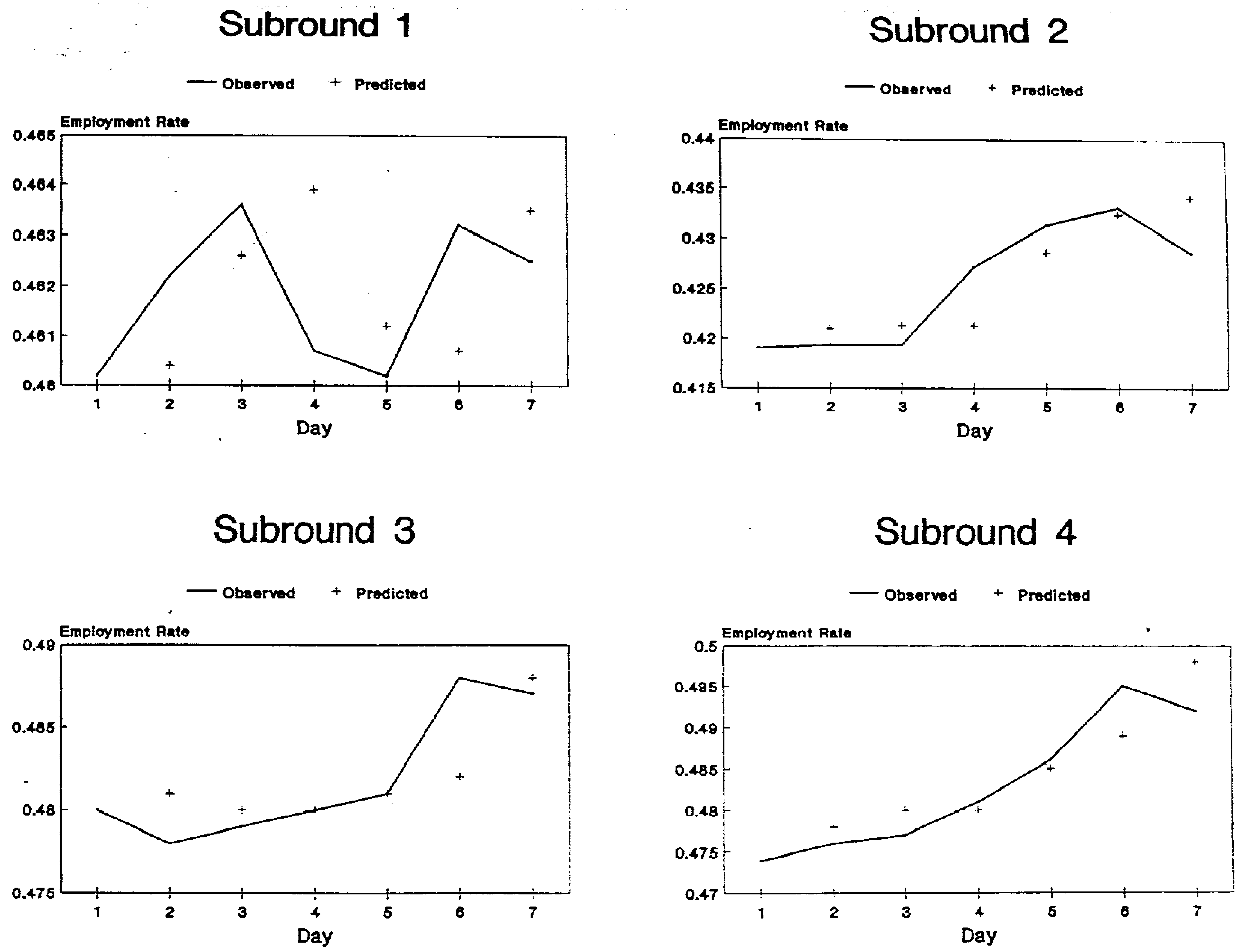

Figure 4.5.1 : Observed and Estimated Daywise Proportion of Employed Persons accross Subrounds: Rural Maharashtra, NSSO 38-th Round, Combined Sample.

(Frydman, 1984; Sampson, 1990), *or* one can assume the transition probabilities to be functions of some important covariates like household occupation, age, sex, education etc. These generalisations will help us in identifying the segment of population who are more shock prone to labour market uncertainties.

Although the empirical problems mentioned above are important and may provide some further insights into the fluctuation of aggregate employment, we do not make any attempt here for such generalisations. In Chapter 5, we shall once again come back to the problem of summarising and measuring employment fluctuation. The results derived in this chapter will be used for this purpose.

Chapter 5

Employment Fluctuation in Rural India : Predictability and Mobility

5.1 Introduction

In Chapter 2 and Chapter 3 we measured employment fluctuation by the total number of state changes observed for individuals in the reference period. The number of jumps from one employment state to another within a specified period basically reflects *the extent of movement* in the corresponding stochastic process. In this chapter a measure of employment fluctuation based on *the extent of predictability* of the time path of the aggregate employment vector is proposed. This measure is derived from the period-to-period variation of the cell aggregates across individuals and has some nice probabilistic and information theoretic interpretations. We apply this measure to the NSSO 38-th round data for rural Maharashtra to determine the extent of employment fluctuation of various subgroups of the population in rural Maharashtra.

This chapter is also a logical follow up of Chapter 4. In Chapter 4 we specified a Markov chain model. The extent of fluctuation of the transition matrix of a Markov chain is often summarised by a scalar. These scalars are called mobility indices. Mobility measures play an important role in summarising the extent of movement and the extent of predictability from a transition matrix. Over the years, many such measures have been proposed and used in economics, sociology and other disciplines (Shorrocks, 1978; Bartholomew 1982; Dardanoni 1993). We shall show that the proposed measure of employment fluctuation can also be interpreted as a measure of mobility over an important class of transition matrices that are frequently encountered in reality.

The plan of this chapter is as follows : Section 5.2 presents a brief survey of the literature on mobility indices. Section 5.3 proposes the new measure and interprets it in different ways. In Section 5.4, the measure is applied to the NSSO 38-th round data on employment to determine the extent of employment fluctuation of various subgroups of the population. Finally, Section 5.5 summarises the main findings and makes some concluding observations.

5.2 A Brief Survey on Measures of Mobility

Studies on mobility either make assumptions directly about the various mobility indicators and analyse their properties (Shorrocks, 1978; Sommers and Conlisk, 1978; Bartholomew, 1982; Geweke *et al.*, 1986; Conlisk, 1990) or focus on the ordering of the transition matrices using some a priori criterion (Markandya, 1984; Conlisk, 1989; Dardanoni, 1993). In this chapter, we shall consider the first approach only.

5.2.1 Criteria for Mobility Indices

An index of mobility is defined as a continuous real valued function $M(\cdot)$ over the set of $K \times K$ transition matrices $\mathcal{P}$. The mobility indices should satisfy the following axioms as specified by Shorrocks (1978) :

(i) **Normalisation (N)** : $0 \leq M(\mathrm{P}) \leq 1$

Note that, normalisation imposes no significant constraint on the set of mobility measures. If $M(\cdot)$ is bounded, a rank preserving change of origin and scale can always be found so that after the change, the transferred function takes values in the interval [0,1].

(ii) **Monotonicity (M)** : Let us define $\mathbf{P} \succ \mathbf{P}'$ if $p_{ij} \geq p'_{ij}$ for all $i \neq j$ and $p_{ij} > p'_{ij}$ for some $i \neq j$. The monotonicity axiom states that $\mathbf{P} \succ \mathbf{P}'$ implies $M(\mathbf{P}) > M(\mathbf{P}')$.

The intuitive implication of this assumption is clear. Monotonicity axiom stresses the importance on *movement*. The extent of movement in the system depends on the values of the diagonal terms of the transition matrix. The lower the diagonal terms, the more is the extent of movement and thus the more mobile the transition matrix is.

(iii.a) **Immobility (I)** : $M(\mathbf{I}) = 0$ where $\mathbf{I}$ is the identity matrix.

The identity matrix arises as a transition matrix when no movement takes place among the states of the Markov chain. If we know the initial distribution, in this case we can accurately predict the time path. The assumption **I** implies that mobility for identity matrix should be less than or equal to mobility for all other transition matrices.

This criterion may be replaced by a stronger criterion, viz.,

(iii.b) **Strong Immobility (SI)** : $M(\mathbf{P}) = 0$ if and only if $\mathbf{P} = \mathbf{I}$.

SI rules out the possibility that besides the identity matrix, any other matrix can take the lowest value.

Although the identity matrix can be associated with the minimum value of a mobility measure, it is harder to decide on the other end point of a mobility measure. In *intergenerational mobility studies* an ideal condition will be one in which the son's class will be independent of his father's class (*origin independence*). Following this idea, Prais (1955) defined a perfectly mobile transition matrix as one all the rows of which are identical.

(iv.a) **Perfect Mobility (PM)** : $M(\mathbf{P}) = 1$ if $\mathbf{P} = \mathbf{u}\mathbf{x}'$ where $\mathbf{u} = (1, 1, \ldots, 1)$ and $\mathbf{x}'\mathbf{u} = 1$, i.e., $M(\mathbf{P})$ is maximum when the transition matrix $\mathbf{P}$ has identical rows.

PM can also be replaced by a very similar criterion, viz., (iv.b) **Strong Perfect Mobility (SPM)** : $M(\mathbf{P}) = 1$ if and only if $\mathbf{P} = \mathbf{u}\mathbf{x}'$ where $\mathbf{u} = (1, 1, \ldots, 1)$ and $\mathbf{x}'\mathbf{u} = 1$.

Like **SI**, **SPM** also rules out the possibility of other type of matrices taking the extreme values.

Note that **PM** or **SPM** specifies an ideal norm. It simply states that mobility should be maximum for the case where the current state (or origin) of an individual does not matter. Since the rows of the transition matrix are identical in this case, the probability of the state in which an individual may go is independent of the current state of the individual and mobility according to **PM** (or **SPM**) should be maximum for this case. Note that although it is a desirable norm with which to compare existing societies in intergenerational problems, it neither conveys maximum amount of movement which can take place between classes nor does it convey the maximum uncertainty in the sense of statistical prediction. In any study of mobility, we must therefore decide whether we are primarily interested in movement, origin independence or in predictability.[1]

We next examine whether meaningful mobility comparisons is possible when the observation periods corresponding to the matrices are not identical. Such a case arises when we have to rank income mobilities of two countries, say North and South, transition matrices for which are obtained by a cross-tabulation of consecutive years' data and data for years separated by a longer period respectively.

Shorrocks(1978) has shown that for many commonly accepted mobility measures $M(\cdot)$, $M(\mathbf{P}) > M(\mathbf{Q})$ does not necessarily imply $M(\mathbf{P}^s) > M(\mathbf{Q}^s)$ for all integer $s \geq 1$. It is this inconsistency which prevents an unambiguous ranking of transition matrices defined over different time intervals. Shorrocks, therefore, defined another criterion called period consistency which is as follows :

(v.a) **Period Consistency (PC)** : $M(\mathbf{P}) > M(\mathbf{Q})$ implies $M(\mathbf{P}^s) > M(\mathbf{Q}^s)$ for all integer $s \geq 1$.

This property is sometimes changed by explicitly introducing into the mobility index $M(\cdot)$ the time interval S over which the transition ma-

[1]Shorrocks (1978) has shown that in general the assumptions **M** and **PM** are incompatible. In Section 5.3, we shall discuss this issue in details.

trix is defined. Thus, a stronger version of **PC** is :

(v.b) **Period Invariance (PI)** : $M(\mathbf{P}, S) = M(\mathbf{P}^s, sS)$.

We shall now show that **PI** relates the mobility index of the transition matrix with its characteristic roots. To see this, let us assume that $\mathbf{P}$ is diagonalisable with distinct characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_K$. Under the usual topology, matrices with distinct characteristic roots form a dense subset within the set of all matrices. Without loss of generality, we assume[2]

$$1 = |\lambda_1| > |\lambda_2| > \cdots > |\lambda_K| \tag{5.2.1}$$

Note that $|\lambda_k| < 1$ because of the assumption that the transition matrix $\mathbf{P}$ admits a steady state solution. We now apply the standard decomposition of $\mathbf{P}$, viz.,

$$\mathbf{P} = \sum_{k=1}^{K} \lambda_k \mathbf{E_k} \tag{5.2.2}$$

where

$$\mathbf{E_k}\mathbf{E_{k'}} = \begin{cases} \mathbf{0} & \text{if } k \neq k' \\ \mathbf{E_k} & \text{if } k = k' \end{cases} \tag{5.2.3}$$

and

$$\sum_{k=1}^{K} \mathbf{E_k} = \mathbf{I} \tag{5.2.4}$$

Note that the matrices $\mathbf{E_k}$ are formed by the product of the right column eigenvector $\mathbf{x_k}$ and the left row eigenvector $\mathbf{y'_k}$ corresponding to λ_k. Under normalisation, $\mathbf{y'_k}\mathbf{x_k} = 1$. Also note that such a decomposition is always possible because of the assumption of distinct characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_K$.

Note that the transition matrix can be written as

$$\mathbf{P}^s = \sum_{k=1}^{K} \lambda_k^s \mathbf{E_k} \tag{5.2.5}$$

To prove **PI**, we must write $M(\mathbf{P}, S)$ as an equivalent function

$$\begin{aligned} M(\mathbf{P}, S) &= f(S, \lambda_2, \ldots, \lambda_K, \mathbf{E_1}, \mathbf{E_2}, \ldots, \mathbf{E_K}) \\ &= f(S, \lambda_2^s, \ldots, \lambda_K^s, \mathbf{E_1}, \mathbf{E_2}, \ldots, \mathbf{E_K}) \end{aligned} \tag{5.2.6}$$

[2]See Shorrocks (1978).

In addition, we shall have to show that the index is an even function of each eigenvalue argument and is homogeneous of degree zero in the time period and logarithms of the characteristic roots of the transition matrix. Both **PC** and **PI** are strong conditions. Very few commonly used mobility measures actually satisfy these criteria.

Geweke *et al.* (1986) showed that Shorrocks's criteria can be logically grouped into three categories, viz., persistence criteria, convergence criteria and temporal aggregation criteria. They showed that criteria within each category is logically consistent. However, across categories several conflicts arise. For matrices with real and nonnegative eigenvalues there are no logical conflicts among any of the criteria. Geweke *et al.* (1986) showed that some of these inconsistencies can be removed if the analysis is extended to continuous time. Their continuous time fractile analysis leads to a one dimensional mobility function of order K, intermediate between a mobility index (dimension zero) and the transition matrix itself (dimension two).

5.2.2 Some Commonly Used Mobility Indices

This subsection presents some commonly used mobility indices and discusses their properties.

- **Measures Based on Trace :** Prais(1955) showed that the mean length of stay in state k is $1/(1-p_{kk})$. Shorrocks(1978) proposed the inverse of the harmonic mean of these lengths, scaled by $K/(K-1)$, as an index of mobility, i.e.,

$$M_T(\mathbf{P}) = \frac{K - \text{ trace}(\mathbf{P})}{K-1} \tag{5.2.7}$$

This measure is simple and is easily interpretable. However, it pays no attention to the off-diagonal elements of $\mathbf{P}$ (Sommers and Conlisk, 1978).

One may define a similar type of mobility index based on the unconditional probability of leaving the current state (Geweke *et al.*, 1986),

$$M_U(\mathbf{P}) = \frac{K}{K-1}\sum_k \pi_k(1-p_{kk}) \tag{5.2.8}$$

Suppose the chain is in steady state. Then $M_U(\mathbf{P})$ is proportional to the probability that a randomly chosen individual will change state in the next period.

- **Measures Based on Determinant** : A mobility measure based on the determinant of a transition matrix in Shorrocks(1978) is

$$M_D(\mathbf{P}) = 1- \mid \det(\mathbf{P}) \mid^{\alpha} \tag{5.2.9}$$

This measure has the obvious weakness that it gives completely mobile value when any two rows or columns of $\mathbf{P}$ are identical (Shorrocks, 1978).

- **Measures Based on Eigenvalue** : Any function $M(\mathbf{P})$ which can be expressed as a strictly decreasing function of the moduli of the eigenvalues of $\mathbf{P}$ can be defined as a mobility index after proper normalisation. Geweke *et al.*(1986) proposed one such index given by

$$M_E(\mathbf{P}) = \frac{K - \sum_k \mid \lambda_k \mid}{K-1} \tag{5.2.10}$$

Measures based on the second highest eigenvalue of the transition matrix have also been proposed (Sommers and Conlisk, 1979), viz.,

$$M_{SC}(\mathbf{P}) = 1- \mid \lambda_2 \mid \tag{5.2.11}$$

where λ_2 is the second highest eigenvalue of $\mathbf{P}$.

Measures based on the second highest eigenvalue are often interpreted in terms of speed of convergence to the steady state solution from a given state, e.g., Shorrocks's measure

$$M_S(\mathbf{P}, S) = e^{-hS} \tag{5.2.12}$$

where $h = -\frac{\log 2}{\log|\lambda_2|}$ and S is the time interval after which transitions are observed, is interpreted as the asymptotic half-life of convergence to the steady state (Shorrocks, 1978).

Note that the mobility indices based on the determinants of the transition matrix can also be interpreted as indices based on eigenvalues because determinants are products of the respective eigenvalues.

- **Bartholomew's Measure :** Bartholomew's mobility index is defined as

$$M_B(\mathbf{P}) = \frac{1}{n-1} \sum_i \sum_j \mid i - j \mid \pi_i p_{ij} \qquad (5.2.13)$$

Here, $M_B(\mathbf{P})$ is proportional to the expected number of class boundaries crossed from one time to the next when the chain is in the steady state.

- **Measures Based on Mean First Passage Time (MFPT) :** Consider a steady state Markov chain. Let two individuals from the population be chosen at random. Conlisk (1990) considers an immobility measure which is the expected number of time periods which must pass before the first individual achieves the state of the second individual. Let $\mathbf{S}^*$ denote the mean first passage matrix (Kemeny and Snell, 1976). The immobility measure is defined as

$$M_P(\mathbf{P}) = \pi / S^* \pi \qquad (5.2.14)$$

- **Measures Based on Entropy or Information :** Besides the above measures, some other measures have also been suggested. Maasoumi and Zandvakili (1986, 1990) applied some generalised entropy measures of mobility to examine the extent of income mobility in a population. Chakravarty (1994) has interpreted the minimum discrimination information statistic suggested by Kullback (1959) as a measure of mobility.

5.2.3 Mobility Indices : A Brief Discussion

Although mobility indices play an important role in summarising the extent of fluctuation, considerable problems are encountered in trying to summarise a transition matrix into a scalar measure of mobility. The indices, ideally, should not only satisfy all the criteria proposed but also have clear interpretations. Very few mobility indices pass through these tests. Moreover, different mobility indices impose different orderings and often these orderings are not consistent with one another. To elucidate this, we present below an illustrative example given in Dardanoni (1993) :

• **Example (Dardanoni, 1993)** : Let us consider the following three transition matrices $\mathbf{P_1}$, $\mathbf{P_2}$ and $\mathbf{P_3}$.

$$\mathbf{P}_1 = \begin{pmatrix} 0.60 & 0.35 & 0.05 \\ 0.35 & 0.40 & 0.25 \\ 0.05 & 0.25 & 0.70 \end{pmatrix} \quad \mathbf{P}_2 = \begin{pmatrix} 0.60 & 0.30 & 0.10 \\ 0.30 & 0.50 & 0.20 \\ 0.10 & 0.20 & 0.70 \end{pmatrix} \quad \mathbf{P}_3 = \begin{pmatrix} 0.60 & 0.40 & 0.00 \\ 0.30 & 0.40 & 0.30 \\ 0.10 & 0.20 & 0.70 \end{pmatrix}.$$

Dardanoni calculated some commonly used mobility indices for $\mathbf{P_1}$, $\mathbf{P_2}$ and $\mathbf{P_3}$ and ordered these three matrices according to each such measure. The results are presented in Table 5.2.1.

Table 5.2.1: Most Mobile Matrix among $\mathbf{P_1}$, $\mathbf{P_2}$ and $\mathbf{P_3}$ for Some Common Mobility Measures

	Second Highest Eigenvalue	Trace	Determinant	Mean First Passage Time	Bartholomew
Most Mobile Matrix	$\mathbf{P_2}$	$\mathbf{P_1}, \mathbf{P_3}$	$\mathbf{P_1}$	$\mathbf{P_3}$	$\mathbf{P_1}, \mathbf{P_2}, \mathbf{P_3}$

From Table 5.2.1, it may be noted that the ranking induced by the measures are so different that one may have to conclude that "no single mobility statistic has the minimum requirements regarded as essential" (Shorrocks, 1978).

5.3 The Proposed Mobility Index $M^*(\mathbf{P})$

In Section 5.2, we provided a brief discussion on some commonly used mobility measures. Note that mobility means not only movement but also predictability. The axiom **M** is a necessary property for the mobility indices aiming to measure total movement and the axiom **PM** is a necessary property for those which aim to measure origin independence.

Shorrocks (1978) has shown that in general the assumptions **M** and **PM** are incompatible. The conflict can be resolved in a number of ways. One obvious way out is to look for different axiomatic structures. The other is to restrict attention to a smaller class of transition matrices $\mathcal{P}^*$ where $\mathcal{P}^*$ is a subset of $\mathcal{P}$.

In many applications, the higher values of the transition probabilities cluster about the main diagonal. However, the assumption that the transition matrix has a *dominant diagonal*, with $p_{kk} \geq \frac{1}{2}$, is too strict. A larger class is obtained by specifying the requirement that the probability of remaining in the same state is no less than that of moving to any other particular state. Formally, if $p_{ii} \geq p_{ij}$ $(i, j = 1, 2, \ldots, K)$, we shall say that the matrix $\mathbf{P}$ has a *maximal diagonal*. A still larger class is the class of *quasi-maximal diagonal* transition matrices for which there exists positive $\mu_1, \mu_2, \ldots, \mu_K$ such that $\mu_i p_{ii} \geq \mu_j p_{ij}$ for all i, j. Let us denote the class of transition matrices having a maximal diagonal by $\mathcal{P}^*$ and that of the quasi-maximal transition matrices by $\mathcal{P}^{**}$. Restricting the attention to $\mathcal{P}^*$ or $\mathcal{P}^{**}$ has one important advantage, viz., $\mathbf{M}$ and $\mathbf{PM}$ are no longer incompatible.

Shorrock's mobility measure based on trace satisfies $\mathbf{M}$ and $\mathbf{PM}$ over the set of transition matrices with *quasi-maximal diagonals*. Although this class is large, Shorrock's measure pays attentions to only the diagonal elements of the transition matrix. Consider i, j, k such that $i \neq j$ and $i \neq k$. If p_{ij} decreases and p_{ik} increases by the same amount δ , Shorrock's measure for mobility will not be affected. Moreover, according to Shorrock's measure, there can be many matrices which are perfectly mobile.

We illustrate below with an example. Consider the following two transition matrices $\mathbf{P_1}$ and $\mathbf{P_2}$.

$$\mathbf{P_1} = \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.50 & 0.25 & 0.25 \\ 0.50 & 0.25 & 0.25 \end{pmatrix} \quad \mathbf{P_2} = \begin{pmatrix} 0.50 & 0.25-\delta & 0.25+\delta \\ 0.50 & 0.25-\delta & 0.25+\delta \\ 0.50 & 0.25-\delta & 0.25+\delta \end{pmatrix}$$

Note that both $\mathbf{P_1}$ and $\mathbf{P_2}$ are quasi-maximal diagonal transition matrices and each has identical rows. Therefore, both $\mathbf{P_1}$ and $\mathbf{P_2}$ are origin independent and hence are perfectly mobile. *However, in terms of predictability, the two matrices may not be similar.*

Suppose, we know that on period t the i-th individual is in state k which means we know that $X_{it}^k = 1$. Suppose now we want to predict the state of the individual on period $(t+1)$. Our accuracy of prediction will clearly depend upon the terms $\mathrm{Var}(X_{i(t+1)}^j \mid X_{it}^k = 1)$ $(j = 1, 2, \ldots, K)$. A mobility measure which is a function of these conditional variances will involve not only the diagonal elements, but all elements of the transition matrix.

We propose such a measure here which is based on the dispersion matrix of the cell aggregates derived in Chapter 4 and interpret it in different ways. The elements of the dispersion matrix of the cell aggregates measure the variability of the cell aggregates. Hence, they can be suitably combined to a real number to provide an idea about the predictability of the transition matrix itself. The measure we propose uses only the diagonal elements of the dispersion matrix of the cell aggregates (i.e., it uses only the conditional variances). However, it involves all the elements of the transition matrix.

Let us define

$$M^{*}(\mathbf{P}) = \mathbf{trace}(\sum_{k=1}^{K} Z_t^k \mathbf{U}_k) = \sum_{k=1}^{K} Z_t^k \mathbf{trace}(\mathbf{U}_k), \qquad \mathbf{P} \in \mathcal{P}^{*} \qquad (5.3.1)$$

where $\sum_{k=1}^{K} Z_t^k \mathbf{U}_k$ is the dispersion matrix of the conditional distribution of the cell aggregates described in Chapter 4.

It can be very easily shown that

$$\mathbf{trace}(\mathbf{U}_k) = 1 - \sum_{j=1}^{K} p_{kj}^2 \qquad (5.3.2)$$

Therefore,

$$M^{*}(\mathbf{P}) = 1 - \sum_{k=1}^{K}\sum_{j=1}^{K} Z_t^k p_{kj}^2 = \sum_{k=1}^{K}\sum_{j=1}^{K} Z_t^k (1 - p_{kj}^2) \qquad (5.3.3)$$

We now provide an information theoretic interpretation to $M^{*}(\mathbf{P})$. Consider a discrete probability distribution with probability vector

$$\tilde{q} = (q_1, q_2, \ldots, q_n)$$

. Then the function

$$h_n(\tilde{q}) = 2(1 - \sum_{k=1}^{n} q_k^2) \qquad (5.3.4)$$

may be interpreted as a non-additive measure of entropy of the corresponding probability distribution (Behara, 1990). In fact, for $n = 2$, $h_2(\tilde{q})$ is called *parabolic entropy* as the graph of $h_2(\cdot)$ becomes a parabola in this case (Behara and Nath, 1973). The proposed measure $M^{*}(\cdot)$ is thus an weighted average of the entropies of the probability distributions corresponding to the rows of a transition matrix, the weight vector being the 'current' or the 'spot' distribution. Thus with this measure we may easily compare the predictability or the information in any two transition matrices. Note that if we interpret $M^*(\cdot)$ as a measure for predictability only, then the comparison of predictability can be done for *all* transition matrices, one need not restrict one's attention to smaller classes.

We shall now consider another simple probabilistic interpretation of $M^*(\mathbf{P})$. Consider two individuals A and B. Let $S_t^{(A)}$ and $S_t^{(B)}$ be two random variables which denote the state positions of A and B respectively on the t-th day. Clearly, both $S_t^{(A)}$ and $S_t^{(B)}$ can take values $k = 1, 2, \ldots, K$, where k denotes the k-th state in the state space. Suppose the spot distribution $\widetilde{Z_t}$ is known. Then

$$P\,[S_t^{(A)} = k] \;=\; P\,[S_t^{(B)} = k] \;=\; Z_t^k \qquad (5.3.5)$$

Also, suppose we *know* that on the t-th day, A and B were on the *same* state, i.e.,

$$P\,[S_t^{(A)} = k \mid S_t^{(B)} = k] = 1 \qquad \forall\, k = 1, 2, \ldots, K \qquad (5.3.6)$$

This means,

$$\begin{aligned} P\,[S_t^{(A)} = k, S_t^{(B)} = k'] &= 0 \qquad \text{if } k \neq k' \\ &= Z_t^k \qquad \text{if } k = k' \end{aligned} \qquad (5.3.7)$$

Also note that

$$P\,[S_{t+1}^{(A)} \neq S_{t+1}^{(B)} \mid S_t^{(A)} = S_t^{(B)} = k] = 1 - \sum_{j=1}^{K} p_{kj}^2 \qquad (5.3.8)$$

Therefore,

$$\begin{aligned} P[S_{t+1}^{(A)} \neq S_{t+1}^{(B)}] &= \sum_k \sum_{k'} P[S_{t+1}^{(A)} \neq S_{t+1}^{(B)}, S_t^{(A)} = k, S_t^{(B)} = k'] \\ &= \sum_k P\,[S_{t+1}^{(A)} \neq S_{t+1}^{(B)} \mid S_t^{(A)} = S_t^{(B)} = k] \cdot P\,[S_t^{(A)} = k, S_t^{(B)} = k] \\ &= 1 - \sum_{k=1}^{K} Z_t^k \sum_{j=1}^{K} p_{kj}^2 \end{aligned} \qquad (5.3.9)$$

This means, $M^*(\mathbf{P})$ is the probability that two individuals A and B will be in different states 'tomorrow' given the information that they belong to the same state 'today'.

Note that the minimum value of this probability is zero. For identity matrix $\mathbf{I}$, $\mathbf{M}^*(\mathbf{I}) = 0$. If individuals always remain in their respective states, naturally they will not ever be in different states given that they were in the same state at some point of time. More generally, if all the rows of the transition matrix $\mathbf{P}$ consist of degenerate distributions, the measure will be zero for the same reason.

Also note that the maximum value of this probability cannot be unity as the minimum value of $\sum p_{ij}^2$ is always greater than zero. The maximum value of this probability occurs when all the rows of $\mathbf{P}$ are $(\frac{1}{K}, \frac{1}{K}, \ldots, \frac{1}{K})$. Thus the maximum value of this probability is $(1 - \frac{1}{K})$ which tends to unity as K increases. Thus for proper normalisation, one should divide $M^*(\mathbf{P})$ by $(1 - \frac{1}{K})$.

We shall now provide a mobility interpretation to this measure. Such a task is, however, difficult. Note that the proposed measure will not be a true indicator of mobility for all transition matrices. In fact, it will not satisfy the criteria of $\mathbf{M}$ or $\mathbf{PM}$ over quasi-maximal diagonals. As an illustration, consider the following three quasi-maximal diagonal transition matrices,

$$\mathbf{P_1} = \begin{pmatrix} 5/6 & 1/6 \\ 1/6 & 5/6 \end{pmatrix} \quad \mathbf{P_2} = \begin{pmatrix} 1/6 & 5/6 \\ 1/6 & 5/6 \end{pmatrix} \quad \mathbf{P_3} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

Note that here $M^*(\mathbf{P_1}) = M^*(\mathbf{P_2})$ for any spot distribution. Therefore $\mathbf{M}$ is violated. Also, $M^*(\mathbf{P_2}) < M^*(\mathbf{P_3})$, so that it is also not a satisfactory measure of origin independence over the class of quasi-maximal diagonal transition matrices.

In fact, if all the rows of a transition matrix $\mathbf{P}$ are identical, i.e., $p_{ij} = p_{1j}$ for all i, $j = 1, 2, \ldots, K$, then

$$M^*(\mathbf{P}) = 1 - \sum_{j=1}^{K} p_{1j}^2 \qquad (5.3.10)$$

which is a measure of entropy of the common probability distribution for the rows of the transition matrix (Behara, 1990). Thus the proposed measure also orders origin independent matrices in terms of the predictability in the steady state distribution. The maximum value of the measure is obtained if $p_{ij} = \frac{1}{K}$ for all i, $j = 1, 2, \ldots, K$.

To reconcile the difficulties, note that matrices like $\mathbf{P_2}$ rarely occurs in practice. Therefore, we shall restrict our attention to a smaller class, viz., maximal diagonal transition matrices (i.e., to $\mathcal{P}^*$). Although the class $\mathcal{P}^*$ is not as big as $\mathcal{P}^{**}$, it covers a large number of transition matrices that occur in reality.[3] In fact, in this dissertation, all the transition matrices obtained belong to the class $\mathcal{P}^*$. For example, all the diagonal elements of the estimated transition matrices in Subsection 4.5 of Chapter 4 were close to unity.

We now prove that $M^*(\mathbf{P})$ is a mobility index over $\mathcal{P}^*$.

Proposition 5.3.1 *$M^*(\mathbf{P})$ is a mobility index for all $\mathbf{P} \in \mathcal{P}^*$.*

Proof : Clearly, **N** and **I** are satisfied by $M^*(\mathbf{P})$. Therefore, we will prove only **M** and **PM**.

For **M**, we shall prove that under the maximal diagonal assumption $\sum_{j=1}^{K} p_{kj}^2$ decreases whenever p_{kj}, $j \neq k$ increases at the expense of p_{kk}. Let,

$$S_0 = \sum_{j=1}^{K} p_{kj}^2 \tag{5.3.11}$$

and without loss of generality, suppose p_{12} increases by an amount δ. Since the rows of $\mathbf{P}$ sum to unity, p_{11} decreases by an amount δ. Also, since we are restricting our attention to matrices with maximal diagonals only,

$$p_{11} - \delta \geq p_{12} + \delta \tag{5.3.12}$$

Let us now define

$$S_1 = (p_{11} - \delta)^2 + (p_{12} + \delta)^2 + p_{13}^2 + \cdots + p_{1K}^2 \tag{5.3.13}$$

Subtracting S_0 from S_1, we obtain

$$S_1 - S_0 = 2\delta(p_{12} + \delta - p_{11}) \tag{5.3.14}$$

Clearly, from (5.3.12), $S_1 < S_0$. A change in any off-diagonal element (at the expense of the corresponding diagonal element) thus increases $\mathbf{M}^*(\cdot)$.

[3]For example, see Shorrocks(1976) and Sampson(1990). All the estimated transition matrices in those articles satisfy the maximal diagonal property.

For **PM**, consider any row of the transition matrix **P**, say the k-th one. Since **P** is a transition matrix,

$$\sum_{j=1}^{K} p_{jk} = 1 \tag{5.3.15}$$

Note that (5.3.15) implies that $\sum_{j=1}^{K} p_{kj}^2$ will be globally minimum when $p_{kj} = \frac{1}{K}$ for all j between 1 and K. Since $\mathbf{P} \in \mathcal{P}^*$, the only case in which all the rows of **P** will be identical is the one having each row of **P** as

$$p_k' = (\frac{1}{K}, \frac{1}{K}, \cdots, \frac{1}{K}) \tag{5.3.16}$$

Since, $\sum_{j=1}^{K} p_{kj}^2$ will be minimum in this case, $M^*(\mathbf{P})$ will be maximum. ■

Note that a feature of the proposed measure is that the value of $M^*(\mathbf{P})$ changes as the weights Z_t^k change. Here, in addition to the transition probabilities, the existing probability distribution or the actual 'spot' distribution enters into the mobility measure $M^*(\mathbf{P})$ as weights corresponding to different rows of the transition matrix. Therefore, it is not possible to compare the mobilities of two matrices **P** and **Q** unless the corresponding spot distributions are also known. Further, here it is possible for the same transition matrix **P** to give rise to different degrees of mobility over time as the spot distribution may change over time.

If in (5.3.3) the Z_t^k's are replaced by another set of fixed weights w_k's, then $M^*(\mathbf{P})$ will still be an indicator of mobility. Once again, the measure can be interpreted as a weighted average of the entropies of the rows of the transition matrix. However, if this is done, the resulting measure can no longer be interpreted as a measure proportional to the trace of the dispersion matrix of the conditional variance of the cell aggregates.

5.4 Empirical Findings

We now apply the proposed mobility index to rank employment fluctuation of various subgroups of the population under consideration. As in Chapters 2 and 3, we shall consider here the same covariates viz., region, occupational category, social and religious class of the households and age, sex and education of the individuals. Since many of these covariates may have important interaction effects with agricultural season (i.e., subround), we present mobility indices separately for each covariate across the four subrounds. In order to have an idea of the extent of statistical error of the estimates, mobility indices corresponding to the subsamples are also calculated. Tables 5.4.1 to 5.4.3 present these estimates.

Although the mobility index computed here uses the spot probability distributions $\widetilde{Z}_t$, the value of the mobility index is not expected to change much from day to day during the reference week. Therefore, we have calculated the index by using $\bar{Z}$ as the weight vector where

$$\bar{Z} = \frac{1}{T}\sum_{1}^{T}\widetilde{Z}_t \tag{5.4.1}$$

Note that if T is large, $\bar{Z}$ can be used as an approximation for the steady state probabilities.

The empirical results presented in this chapter are supposed to reveal a clearer picture of the nature and the extent of employment fluctuation than what has been discussed in Chapter 2. Since mobility across all the covariates are calculated separately by subrounds, the nature of the interaction of these covariates with agricultural seasons can be easily examined. In fact, the results obtained suggest some interaction of household occupation and sex with season. This is one significant improvement in our understanding of the nature of employment fluctuation over the inferences drawn in Chapter 2.

The regionwise results suggest that the Northern and the Inland Eastern (Ineast) regions have consistently higher mobility (i.e., greater employment fluctuations) compared to the other four regions of the State's rural sector. This pattern of regional variation is seen to be more or less stable over the subrounds.

Across occupations, the mobility of the wage dependent households are far more than the rest of the population. Among them, it is the agricultural labourers who are the most affected ones. The results suggest that during the period July to September, their mobility can be as high as 0.1355, while the indices corresponding to all other occupational groups are less than 0.05 for that period. However, note that during the period April to June, which is the lean season, the employment fluctuations of the agricultural labourers are less. Since the rate of unemployment is high in the lean season, this means the *duration of unemployment* during the lean season is higher.

Across social groups, mobility is high for SC and ST persons – their mobility being approximately double that of other Hindus for all the subrounds. However, note that the mobility of the category Othgrp is also quite high. This is because the category Othgrp contains the Neo-Buddhists who are converted SC and ST persons. Since they are mostly poor and do not have an adequate asset base as a protection against unemployment, they probably suffer from frequent spells of underemployment and hence their employment fluctuation is also high.

Across demographic categories, employment fluctuation is low for children and old persons. This is because children and old persons mostly remain out of labour force and rarely change their status. This is clearly consistent with Chapter 2. In Chapter 2 we also noticed that the average number of jumps for girl children are higher than that for boy children. The seasonal breakup, however, reveals that mobility of girls become slightly less in the lean season, i.e. from April to June. While the mobility of the young boys are 0.0196 in that period, that of the girls are only 0.0096. This is not only true for the combined sample, but also separately true for both the subsamples. However, in all the other seasons, employment fluctuation are higher for girls. The seasonal breakup also provides clearer pictures of the mobilities of adult females vis-a-vis adult males and old females vis-a-vis old males. Employment fluctuation for women is low in Subround 2, being only 0.0796 compared to 0.1042 for the adult males. However, in all other subrounds, mobility of women is more, possibly because they move more frequently in and out of labour force – the difference being more pronounced in Subround 3. Mobility of the old persons are low. Among old persons, however, mobility of males are more than that of females.

Across education, the seasonal break up yields the same type of results as has been found in Chapter 2.[4] Mobility of the illiterates are higher than those who have primary education. The highly educated persons, on the other hand, demonstrate little employment fluctuation. The measure of mobility for them varies between 0.0212 to 0.0275 only.

Thus, the above analysis suggests that among the covariates considered, *households' main occupational type* and *age and sex* have some interac-

[4]We have already interpreted the results in details in Chapter 2.

Table 5.4.1: Mobility of Various Covariate Groups Separately across Subrounds : Rural Maharashtra, NSSO 38-th Round, Subsample 1

Covariates	Categories	Subround 1	Subround 2	Subround 3	Subround 4
(1)	(2)	(3)	(4)	(5)	(6)
Region	Coastal	0.0307	0.0336	0.0242	0.0358
	Western	0.0256	0.0561	0.0487	0.0759
	Northern	0.1360	0.0823	0.0807	0.0841
	Central	0.0470	0.0738	0.0327	0.0553
	Ineast	0.0894	0.0887	0.1252	0.0823
	Eastern	0.0592	0.0237	0.0935	0.0574
Household Type	Selfagr	0.0256	0.0332	0.0316	0.0352
	Selfnagr	0.0290	0.0197	0.0558	0.0209
	Agrlab	0.1134	0.1051	0.1238	0.1073
	Othlab	0.0537	0.0823	0.0523	0.1076
	Othocc	0.0119	0.0227	0.0276	0.0275
Social Group	ST	0.0898	0.1096	0.0970	0.0954
	SC	0.0771	0.0908	0.0813	0.0934
	Othhind	0.0425	0.0432	0.0518	0.0584
	Othmusm	0.0591	0.0277	0.1110	0.0477
	Othgrp	0.0925	0.1625	0.1021	0.0824
Age and Sex	Boychild	0.0043	0.0162	0.0053	0.0113
	Girlchild	0.0190	0.0075	0.0403	0.0307
	Adultmale	0.0795	0.0933	0.0796	0.0866
	Adultfemale	0.0902	0.0878	0.1106	0.1063
	Oldmale	0.0438	0.0597	0.0498	0.0349
	Oldfemale	0.0173	0.0263	0.0163	0.0010
Education	Illit	0.0771	0.0809	0.0899	0.0897
	Primid	0.0410	0.0418	0.0451	0.0486
	Highed	0.0247	0.0158	0.0188	0.0192

tions with agricultural seasons. The other covariates may not have such a strong interaction with seasons.

Table 5.4.2: Mobility of Various Covariate Groups Separately across Subrounds : Rural Maharashtra, NSSO 38-th Round, Subsample 2

Covariates	Categories	Subround 1	Subround 2	Subround 3	Subround 4
(1)	(2)	(3)	(4)	(5)	(6)
Region	Coastal	0.0126	0.0168	0.0364	0.0461
	Western	0.0464	0.0624	0.0865	0.0543
	Northern	0.0868	0.0709	0.1003	0.0953
	Central	0.0498	0.0442	0.0779	0.0551
	Incast	0.0704	0.0937	0.1123	0.0683
	Eastern	0.0533	0.1046	0.0357	0.0725
Household Type	Selfagr	0.0292	0.0407	0.0390	0.0287
	Selfnagr	0.0289	0.0396	0.0411	0.0471
	Agrlab	0.0932	0.0987	0.1451	0.1096
	Othlab	0.1112	0.0920	0.0367	0.0847
	Othocc	0.0180	0.0290	0.0409	0.0221
Social Group	ST	0.0707	0.1110	0.1164	0.0863
	SC	0.0496	0.0977	0.1398	0.1232
	Othhind	0.0495	0.0543	0.0662	0.0493
	Othmusm	0.0747	0.0459	0.0668	0.0529
	Othgrp	0.0520	0.0576	0.0815	0.0688
Age and Sex	Boychild	0.0069	0.0174	0.0153	0.0039
	Girlchild	0.0276	0.0116	0.0308	0.0322
	Adultmale	0.0780	0.1146	0.1069	0.0753
	Adultfemale	0.0732	0.0710	0.1289	0.1007
	Oldmale	0.0342	0.0881	0.0271	0.0278
	Oldfemale	0.0171	0.0151	0.0223	0.0593
Education	Illit	0.0689	0.0731	0.1038	0.0848
	Primid	0.0351	0.0567	0.0596	0.0370
	Highed	0.0305	0.0245	0.0315	0.0274

The reliability of the above estimates of mobility can be verified from the subsamples. In almost all the cases, the estimates from the combined sample was found to lie in between the corresponding estimates from the subsamples, thus confirming the robustness of the results. However, for many covariate categories, the distance between the estimates were high. This, however, somewhat weakens the results obtained.

Table 5.4.3: Mobility of Various Covariate Groups Separately across Subrounds : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Covariates	Categories	Subround 1	Subround 2	Subround 3	Subround 4
(1)	(2)	(3)	(4)	(5)	(6)
Region	Coastal	0.0215	0.0262	0.0327	0.0411
	Western	0.0358	0.0597	0.0699	0.0668
	Northern	0.1112	0.0773	0.0925	0.0904
	Central	0.0485	0.0582	0.0548	0.0555
	Ineast	0.0801	0.0913	0.1200	0.0759
	Eastern	0.0563	0.0664	0.0703	0.0650
Household Type	Selfagr	0.0272	0.0368	0.0349	0.0320
	Selfnagr	0.0289	0.0306	0.0478	0.0344
	Agrlab	0.1006	0.1024	0.1355	0.1084
	Othlab	0.0821	0.0882	0.0445	0.1004
	Othocc	0.0162	0.0274	0.0356	0.0258
Social Group	ST	0.0829	0.1103	0.1086	0.0911
	SC	0.0617	0.0940	0.1203	0.1081
	Othhind	0.0463	0.0487	0.0591	0.0540
	Othmusm	0.0696	0.0386	0.0890	0.0540
	Othgrp	0.0751	0.0900	0.0932	0.0778
Age and Sex	Boychild	0.0056	0.0169	0.0106	0.0076
	Girlchild	0.0235	0.0096	0.0357	0.0317
	Adultmale	0.0787	0.1042	0.0942	0.0812
	Adultfemale	0.0819	0.0796	0.1203	0.1036
	Oldmale	0.0388	0.0760	0.0416	0.0323
	Oldfemale	0.0176	0.0205	0.0211	0.0314
Education	Illit	0.0728	0.0772	0.0974	0.0872
	Primid	0.0381	0.0494	0.0523	0.0432
	Highed	0.0275	0.0212	0.0255	0.0232

5.5 Conclusion

In this chapter we have proposed a measure of employment fluctuation based on the notion of predictability of the time path of the aggregate employment. The measure we have proposed is proportional to the trace of the dispersion matrix of the conditional distribution of the state aggregates and has some simple probabilistic and information theoretic interpretations. It is a weighted average of entropies of the probability distributions corresponding to the rows of the transition matrix. Also, it is proportional to the probability that the two individuals A and B will be in different states on the next period, given the information that they are in the same state in the current period. We have also shown that the proposed measure can also be interpreted as a measure for mobility for the class of maximal diagonal transition matrices.

We have used this measure to determine the extent of employment fluctuation of various subgroups of the population in rural Maharashtra. The empirical exercise revealed that among the regions, employment fluctuates more in the Northern, Inland Eastern and the Eastern parts of Maharashtra. Also, the agricultural labourers and the SC and the ST persons suffer more from employment fluctuation. Those who are highly educated, on the other hand, enjoy steady employment patterns.

Before concluding this chapter, we shall indicate some limitations of this present study. These limitations are both theoretical and empirical in nature. In mobility studies, we are not only concerned with movement but also with predictability. This chapter attempted to define mobility in terms of predictability of the cell aggregates. The measure proposed in this chapter can be used as a mobility index *only when the transition matrices have maximal diagonals*. Also, the criterion of **PI** is not satisfied by this index.

Empirically, although the ranking implied by this index mostly agrees with the other methods of characterising employment fluctuation described in Chapters 2 and 3, it has some limitations. We have calculated employment mobility of several covariate groups separately across the subrounds. We have not examined the effect of all of the covariates *simultaneously*. Thus the 'true' incremental contribution of the covariates have not been obtained. Also, the measures across the subsamples indicate that the standard errors of some of the estimates are high. Needless to mention that the high standard error considerably weakens the reliability of the results obtained.

Although these problems are important, generalisations in these directions are beyond the scope of this dissertation. In Chapter 6, we shall measure *employment uncertainty* of various subgroups of the population in another way. Chapter 6 examines and generalises some traditional models which explains the variations of unemployment rate over various subgroups of the population. Our method in Chapter 6 will not only provide this *traditional explanation*, it will also enable us to rank those covariate groups according to the extent of employment risk.

Chapter 6

Employment Risk in Rural India

6.1 Introduction

In the earlier chapters we have seen that in the Indian rural sector employment fluctuates considerably even within a very brief reference period. This, in other words, suggests that the households are exposed to the risk of unemployment. One would expect such risk to be affected by the market conditions and also by several village and household specific factors in the rural sector. This chapter examines the extent of fluctuation of employment and tries to identify the socioeconomic and household specific variables responsible for such fluctuation.

The existing literature on the empirical analysis of employment has generally ignored this problem of variation of employment risk across households. In almost all of these studies some aggregate measures of employment is regressed on its determinants, assuming constant variance of the disturbance term across observations (Bardhan, 1984; Paul, 1988). Clearly, this amounts to assuming that the employment risk is same across observations. Apparently, adequate attention has not been given to ascertain the effect of socioeconomic as well as household specific variables on employment risk.

Also, the empirical models for the analysis of employment rate proposed so far are somewhat ad-hoc in the sense that they often lack a micro foundation. In this chapter we shall first specify the employment behaviour of individuals and then consider the aggregation implications of such employment behaviour. To be precise, we shall consider *the aggregation implication* of two intuitively appealing assumptions about employment behaviour, viz., (i) individuals' day-to-day employments are independent *and* (ii) individuals' day-to-day employments follow independent Markov chains.

We shall show that under the above assumptions and certain intuitively appealing specifications, the distribution of some aggregate measures of employment will follow the class of models proposed by Just and Pope(1978, 1979). In these models, the dependent variable is heteroscedastic and its conditional variance is a function of the explanatory variables. Such models have already been used to analyse the pattern of risk associated with the use of various inputs in agricultural production. We shall use these models to test the presence of heteroscedasticity in the employment rate using the data on employment and unemployment collected by the NSSO for the rural parts of the state of Maharashtra in their 38-th round survey operation.

The plan of this chapter is as follows : Section 6.2 proposes a model of aggregate employment rate which is derived through aggregation of individual employment behaviour and captures the phenomenon of employment risk mentioned above. Section 6.3 presents the empirical results of fitting this model to data. Finally, Section 6.4 summarises the main findings and makes some concluding observations.

6.2 The Model

Suppose there are n macro units in the sample. These units may be villages, regions or states depending upon the level of aggregation considered. In this chapter, we shall refer to these units as villages. Let there be n_i individuals from the i-th village of our sample. Thus, $N = \sum_{i=1}^{n} n_i$ denotes the total number of individuals observed in the sample. For simplicity, let us assume that on any given date, an individual may be either employed (E) or unemployed (U). This means, we are excluding the state "out of labour force" (O) from our analytical framework.[1]

Suppose each individual is observed for T successive days. Since on any given day an individual may be either employed or unemployed, we define

[1]Since we are trying to find out the quantitative nature and extent of employment risk, we restrict our attention to only those who are in the labour force.

a set of random variables

$$\begin{aligned} X_{ijt} &= 1 \quad \text{if the } j\text{-th person of the } i\text{-th village} \\ &\qquad\ \text{is employed on the } t\text{-th day} \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{6.2.1}$$

We also define a random vector

$$\widetilde{X_{ij}} = (X_{ij1}, X_{ij2}, \ldots, X_{ijT}) \tag{6.2.2}$$

which gives the record of employment status of the j-th person of the i-th village for T successive days.

With different specifications of the individual behaviour, different types of models of aggregate employment rate may be obtained. Here we shall consider the aggregation implications of two simple and intuitively appealing assumptions, viz., (i) individuals' day-to-day employments are independent *and* (ii) individuals' day-to-day employments follow independent Markov chains.

6.2.1 The Assumption of Independence

Suppose, X_{ijt}'s are independent across i, j and t. Also, let us assume that X_{ijt}'s are identical for fixed i and j (i.e., across time the probability distributions of individuals remain the same). Let us denote

$$P[X_{ijt} = 1] = p_{ij} \tag{6.2.3}$$

and specify p_{ij} to be a linear function of the major individual and household specific variables, i.e.,

$$p_{ij} = \alpha_0 + \sum_k a_k Z_{ijk} \tag{6.2.4}$$

where α's are a set of parameters to be estimated from data and Z_{ijk}'s are a set of individual and household specific variables.

Let us denote the proportion of days an individual is employed by

$$\bar{X}_{ij} = \frac{1}{T}\sum_{t=1}^{T} X_{ijt} \tag{6.2.5}$$

Clearly, according to the flow of time definition of employment $\bar{X}_{ij}$ measures the individual employment rate.

Since for fixed i and j, X_{ijt}'s are independent and identically distributed (*iid*) Bernoulli(p_{ij}) random variables,

$$\bar{X}_{ij} \sim N(p_{ij}, \frac{p_{ij}(1-p_{ij})}{T}) \tag{6.2.6}$$

for large T. Note that even in this simple specification, application of ordinary least squares (OLS) to a regression model which has the employment rate $\bar{X}_{ij}$ as the dependent variable will not be valid as the variance of $\bar{X}_{ij}$ depends on the individual and household specific variables, Z_{ijk}'s, in view of (6.2.4).

The distribution (6.2.6) falls under the general category of models which are specified in the context of agricultural production risk by Just and Pope(1978, 1979). They specified agricultural production Y_i to be related to a set of independent variables $\tilde{V}_i$ as

$$Y_i = g(\tilde{V}_i) + \epsilon\, h(\tilde{V}_i) \tag{6.2.7}$$

where ϵ is the disturbance term in the regression equation.

The simple specification (6.2.1)-(6.2.6) thus provides an example in which the disturbance term in the regression equation does not have a constant variance. Here, the extent of employment risk will depend on a set of household specific variables.

Note that starting from this individual specific assumptions, we can easily find out the probability distribution of various aggregate employment measures like village employment rate or employment rate of a particular region. These measures will have similar types of distributions as in (6.2.6). These aggregate measures are used when data are available only at the aggregate level and they play a major role in policy determination.

6.2.2 The Assumption of Markov Chain

The independence of day-to-day employment of individuals is a restrictive assumption. Therefore, we next consider a slightly more general framework. We shall assume that the day-to-day employment status of individuals follow a Markov chain. The Markov chain assumption is a considerably more realistic one and is quite frequently adopted in studies of labour force dynamics (David,M. and Otsuki, 1968; Denton, 1973).

We assume that for fixed i and j, X_{ijt}'s follow a Markov chain, but across i and j, X_{ijt}'s are independent, i.e., $\widetilde{X_{ij}}$'s are independent random vectors for all i and j.

Let the transition matrix of the Markov chain $\mathbf{P^{ij}}$ be given by

$$\mathbf{P^{ij}} = \begin{pmatrix} p_{EE}^{ij} & 1 - p_{EE}^{ij} \\ 1 - p_{UU}^{ij} & p_{UU}^{ij} \end{pmatrix} \tag{6.2.8}$$

We also assume that for all i and j, the transition matrices $\mathbf{P^{ij}}$'s are recurrent and yield *unique stationary* distributions. Let π_E^{ij} be the stationary probability of employment of the j-th member of the i-th village. We shall assume that these steady state probabilities are linear functions of the covariates, i.e.,

$$\pi_E^{ij} = \beta_0 + \sum_k \beta_k Z_{ijk} \tag{6.2.9}$$

where β's are a new set of parameters to be estimated from data.

Note that in this specification,

$$\bar{X}_{ij} \xrightarrow{\mathcal{L}} \pi_E^{ij} \tag{6.2.10}$$

for large T. Thus, even for a large value of T, $\bar{X}_{ij}$ becomes a Bernoulli(π_E^{ij}) random variable. Hence, not only does the method of OLS become invalid in this case, the method of generalised least squares (GLS) as described in Subsection 6.2.1 will also *not* be applicable for the purpose of inference as the probability distribution is non-normal.

Now, since employment behaviour of individuals are assumed to be independent, by the Central Limit Theorem (CLT) averages across individuals will converge to a normal distribution . However, since the number of members in a household is generally very small, such average employment rates for households cannot be approximated by a normal distribution. We shall, therefore, restrict our attention to more aggregated measures.

Let us define

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \bar{X}_{ij} \tag{6.2.11}$$

as the employment rate for the i-th village.

By CLT, we know that $\bar{X}_i$ converges in distribution to a normal random variable, i.e.,

$$\bar{X}_i \sim N\left(\frac{1}{n_i} \sum_{j=1}^{n_i} \pi_E^{ij}, \frac{1}{n_i} \sum_{j=1}^{n_i} \pi_E^{ij}(1 - \pi_E^{ij})\right) \tag{6.2.12}$$

Note that since π_E^{ij}'s are functions of the socioeconomic and household specific variables, the probability distribution of $\bar{X}_i$ will also be heteroscedastic. Also note that the two different models in Subsection 6.2.1 and Subsection 6.2.2 yield the same type of functional forms for heteroscedasticity, although at different levels of aggregation.

The above discussion thus clearly brings out the following points : (i) just by regression on aggregate measure of employment rate (e.g., the village level employment rate) on a set of explanatory variables will not reveal the stochastic mechanism governing individual employment behaviour, and (ii) the regression model referred to above will be necessarily heteroscedastic. In fact, this element of heteroscedasticity may be interpreted as the risk associated with the employment of individuals belonging to a specified region and can be analysed empirically.

In this chapter, we restrict our attention only to the nature and the extent of employment risk and examine how it changes across various covariates. We do not attempt to find out the underlying stochastic mechanism yielding such a solution.

6.3 Empirical Findings

The variables used in the empirical study are somewhat similar to those used in Chapter 3. Once again, we club the Coastal and the Inland Western region of Maharashtra together to form a single region and use a dummy variable "West" to distinguish this region from the rest of the State. To capture the phenomenon of seasonality, we use three season dummies to distinguish Subround 2, Subround 3 and Subround 4 from Subround 1. To capture the difference across occupational categories, another dummy variable called "Lab" is introduced. When aggregated across individuals in the sample village, this explanatory variable becomes the proportion of labour households in the village (Labpc).

Since access to an adequate asset base is likely to protect an individual against the risk of unemployment, we bring in the amount of per-capita land cultivated by the household as an important covariate. Also, since quality of land may affect the extent of employment of a household, the amount of per-capita land irrigated is taken as another major explanatory variable. In our

Table 6.3.1: Some Descriptive Statistics of the Variables Covered in the Study

Name	Mean (A)	Variance (V)	C.V.[1]	Minimum	Maximum
(1)	(2)	(3)	(4)	(5)	(6)
Empvil	0.9317	0.0069	0.0894	0.5556	1.0000
West	0.4116	0.2426	1.1968	0.0000	1.0000
Subrnd2	0.2575	0.1916	1.6996	0.0000	1.0000
Subrnd3	0.2171	0.1703	1.9009	0.0000	1.0000
Subrnd4	0.2655	0.1954	1.6650	0.0000	1.0000
Labpc	0.4353	0.0506	0.5166	0.0000	1.0000
Landcult	0.8479	0.4867	0.8228	0.0000	5.0000
Irrign	0.1124	0.0370	1.7116	0.0000	2.0341
Villsamp	48.673	133.45	0.2373	1.0000	86.000

[1] C.V. is the Coefficient of Variation, C.V. = $(\sqrt{V}/A) \times 100$

Villsamp represents number of sample persons in the village

actual analysis, these variables are measured as the village-level averages. We denote the variables as Landcult and Irrign and they measure the per-capita land cultivated in the village and per capita land irrigated in the village respectively.

Finally, it should be pointed out that many other covariates which are likely to be important for determining employment rate of individuals (like demographic particulars of individuals), need not be considered in the specification of the aggregate or average employment rate function because when averaged, these variables will have very little variation across villages. Essentially for this reason, we have not considered these demographic variables in the present analysis.

6.3.1 Summary Measures of the Variables

In order to get a preliminary idea about the extent and nature of variation of the village employment rates (Empvil), we present some descriptive measures of the variables under study in Table 6.3.1.

Since in the NSSO 38-th round survey, only ten households were selected from each sample village, average employment rate for the village has been calculated as the average of the employment rates of all the individuals belonging to these ten sample households. On an average, a typical village employment rate was calculated as the average of employment rates of around twenty five to thirty persons. This lack of adequate number of observations corresponding to each village considerably limited the scope of the present empirical investigation.

Figure 6.3.1 presents the empirical relative frequency distribution of village employment rates. The average village employment rate is found to be 0.93058. However, Figure 6.3.1 also reveals that there exists considerable amount of variation in village employment rate. This is somewhat surprising because it is generally believed that the Employment Guarantee Scheme (EGS) introduced in Maharashtra in the early 70's has considerably removed seasonal discrepancies in rural employment in the State. Though much of the seasonal variations is likely to have been removed, there exists some regional discrepancy in the EGS employment (Ezekiel and Stuyt, 1990). The villages which did not enjoy the benefits of EGS employment were more prone to labour market uncertainties (Walker and Ryan, 1990).

6.3.2 Regressions for the Level of Employment

We present the result of regression of Empvil Table 6.3.2. Although the fit is poor, the interpretations of the coefficients are clear. Village employment rate is more or less same throughout the year. It increases by 2% in Subround 4 and this change is highly significant. The regression coefficient of the region dummy West is not significant. Thus, there seems to exist little difference between the village level employment rates for Western and other regions of Maharashtra. The other variables which turn out to be significant in explaining the variations of village employment rate are Labpc and Landcult. Thus, the villages with a greater proportion of wage dependent persons, i.e., villages with a higher Labpc, are more prone to unemployment. This result is fully consistent with that of Sundaram and Tendulkar(1988), who found that one of the most important determinants of person day unemployment rate (PDUR) in a region is the percentage of wage dependent households (WDH). Also, the regression results indicate that employment in villages with high amount of cultivable land is *more* since an adequate

Figure 6.3.1 : Distribution of Village Employment Rate: Rural Maharashtra, NSSO 38-th Round, Combined Sample.

Table 6.3.2: Regression Results Corresponding to Empvil : Rural Maharashtra, 38-th Round, Combined Sample

Explanatory Variables	Regression Coefficients	Standard Error	t-Value	Significance at Percent Level
(1)	(2)	(3)	(4)	(5)
West	0.0213	0.0099	2.1515	0.1385
Subrnd2	-0.0206	0.0084	-2.4524	0.1232
Subrnd3	-0.0066	0.0015	-4.4000	0.0712
Subrnd4	0.0175	0.0002	87.5000	0.0036
Labpc	-0.0784	0.0055	-15.2545	0.0208
Landcult	0.0130	0.0008	16.2500	0.0195
Irrign	0.0367	0.0174	2.1092	0.1409
Constant	0.9441	0.0043	219.5581	0.0014

Number of observations =540
$R^2 = 0.1367, F = 12.039$, (From Mean)
Residual sum=-0.27, Residual Variance=0.006
Sum of Absolute Errors = 31.130
Runs Test : 247 runs, 327 positive, 213 negative
Normal Statistic = -1.0791
Coefficient of Skewness = -1.2638
Coefficient of Excess Kurtosis = 2.0337

asset base in the village protects the village population from labour market shocks. However, the quality of this land is not found to be a significant factor.

6.3.3 Results of Tests for Heteroscedasticity

The summary statistics related to the regression results presented in Table 6.3.2 shows that the average of the absolute values of the residuals is seen to be 0.0556. Considering the fact that there is little variation in the village employment rates, this average should be considered to be rather high. However, as the value of the normal statistic of -1.0791 suggests, the residuals are more or less symmetrically distributed around zero.

Although the nonparametric run test seems to support the hypothesis of normality of residuals, tests on heteroscedasticity reveal that there exists significant differences across various subgroups of population so far as employment risk is concerned. In Table 6.3.3 we present the results of some heteroscedasticity tests all of which point to the presence of significant het-

Table 6.3.3: Results of Various Tests on Heteroscedasticity

Description Of the Test	Obtained Value of Chi-square	Degrees of Freedom	Critical value of Chi-square (5%)
(1)	(2)	(3)	(4)
Breusch-Pagan-Godfrey	29.184	7	
Glesjer	80.474	7	18.475
Harvey	58.859	7	

eroscedasticity. It may be noted that here we consider only those tests of heteroscedasticity which examine whether the variance of the disturbance term changes with some of the explanatory variables in the regression equation.

6.3.4 The Determinants of Heteroscedasticity

We now examine the effect of the explanatory variables on employment risk as reflected in the heteroscedasticity of the village level employment rate. To do so, we estimate a regression equation which has the absolute values of the residuals (Riskvil) as the dependent variable. All the explanatory variables of the regression equation for the village employment rate are used as explanatory variables for heteroscedasticity. A positive association of any of these explanatory variables with the absolute value of the residual will imply that employment rate becomes *less accurately predictable* as the value of the variable increases. This means that the variance of the disturbance term will be positively related with that variable.

For a better explanation of the absolute value of the residuals, we add another explanatory variable, viz., Villsamp. Villsamp is the number of sample individuals in a village. Intuitively it seems that larger the number of persons in a sample village, the more accurate will be the prediction of the employment rate by the explanatory variables. Hence villages with large number of persons will have small residuals. We add the variable Villsamp to tackle this heterogeneity due to difference in sample size across villages. Table 6.3.4 presents the results of the regression analysis examining the nature of employment risk.

From Table 6.3.4 we see that employment fluctuation in a village is

Table 6.3.4: Regression Results Corresponding to Riskvil : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Explanatory Variables	Regression Coefficients	Standard Error	t-Value	Significance at Percent Level
(1)	(2)	(3)	(4)	(5)
West	-0.0103	0.0132	-0.7803	0.2890
Subrnd2	0.0194	0.0110	1.7636	0.1640
Subrnd3	0.0053	0.0003	17.6667	0.0178
Subrnd4	-0.0009	0.0068	-0.1324	0.4581
Labpc	0.0511	0.0085	6.01118	0.0517
Landcult	-0.0079	0.0034	-2.3235	0.1294
Irrign	-0.0195	0.0075	-2.6351	0.1156
Villsamp	-0.0008	0.0001	-8.0000	0.0396
Constant	0.0839	0.0025	33.5600	0.0086

Number of Observations = 540
$R^2 = 0.1588, F = 12.531$ (From Mean)
Residual Sum = -0.48, Residual Variance = 0.002
Sum of Absolute Errors = 17.336
Runs Test : 259 Runs, 213 positive, 327 negative
Normal Statistic= 0.0030
Coefficient of Skewness = 1.9585
Coefficient of Excess Kurtosis = 5.6360

again dependent on percentage of wage dependent households in the village. Irrigation, as is to be expected, decreases employment fluctuation. However, its overall contribution is not significant. Employment fluctuation is found to increase significantly in Subround 3. However, we obtained no significant differences in employment risk across geographical regions of Maharashtra.

The absolute values of the residuals decrease slightly as the number of samples from the village increases. Although this difference is small, it is significant. This seems to imply that employment risk gets more evenly spread in villages with larger population.

6.4 Conclusion

This chapter has examined the importance of employment risk theoretically as well as empirically for the Indian rural sector. It has been shown that when there exists substantial heterogeneity in the population, classical linear regression models in which variations of some aggregate measures of employment are explained in terms of other covariates may not be an appropriate specification. Starting from some intuitively appealing individual specific assumptions, we have shown here that the distribution of aggregate employment rate will be heteroscedastic.

The empirical exercises done with the help of NSSO 38-th round data on employment and unemployment for the rural sector of the state of Maharashtra clearly reveals that the level of employment is considerably more for the households which are self-employed with an adequate asset base like land. Further, all the tests for heteroscedasticity rejected the assumption of constant variance of employment rate. Among the various subgroups of the population of Maharashtra, the labour households are seen to be exposed to considerably larger employment risk than others. The other covariates like region, social groups etc. are seen to be less important so far as the difference in employment risk is concerned.

We now discuss some limitations and potential sources of generalisations of this model. First, it may be noted that the observed value of the dependent variable, viz., the village employment rate, is always bounded by zero and one. Under the proposed models, the estimated value of this variable, however, may not obey this restriction. Imposition of a parametric restriction to guarantee that the estimated values fall within the bounds will make the process of estimation considerably more complex. However, this is one potential source of generalisation that may yield some more interesting conclusions.

Second, the model presented in this chapter has been derived under the assumption that the day-to-day employment of individuals follow independent, stationary and recurrent Markov chains. This assumption, though intuitively appealing, may not be necessarily true. Even if it is true, convergence to stationarity, once the process is disturbed from the steady state, may take some time. Therefore, the employment rate, calculated on the basis of information on employment status of individuals for a reference period of seven days only, may indeed be a very crude approximation to the true steady state probability of employment. If convergence to stationarity takes more time, the disturbance term in the regression equation will significantly differ from that of a normal distribution.

Finally, a major generalisation of the model may be made by increasing the number of employment states. This can be done either by considering various types of employment (like farm employment, non-farm employment etc.) or by introducing the state 'out of labour force' in the model. The resulting model will be a multivariate one. Though the estimation procedure will be far more complex, such an analysis may help us to provide a micro-theoretic basis for many more single equation and simultaneous equations studies (Dev, 1990; Unni, 1991; Shukla, 1991).

Chapter 7

Conclusion

All the previous chapters of this dissertation together have provided a somewhat detailed discussion on the problem of employment fluctuation. This chapter summarises the main findings and discusses some limitations and potential generalisations of our study. Much of what we say in this chapter has already been discussed in the concluding sections of the preceding chapters. This chapter merely puts them together to provide a more concise and unified picture of the whole problem.

The dissertation attempts to find out methods for studying the nature and the extent of employment fluctuation in rural India based on the available household level data on employment and unemployment. Thus, the dissertation is partly theoretical and partly empirical in nature. It proposes some new methods for *measuring* or *ordering* employment fluctuation and then illustrates the use of these new analytical tools. The empirical analyses presented in this dissertation are based on the household level data on employment and unemployment for rural Maharashtra collected by the National Sample Survey Organisation, Government of India, in their 38-th round survey operation.

On the methodological side, we have tried to build up a consistent analytical framework. Employment fluctuation may be measured in a number of alternative ways. Broadly, there are two aspects of such a measure, viz., the extent of movement and the extent of predictability. In our study, we have focussed on both of these aspects using some traditional analytical tools and proposing some new ones.

Chapters 2 and 3 of this dissertation are concerned with the aspect of movement. In these chapters, we have measured employment fluctuation by the number of state changes committed by individuals. A preliminary data analysis has been carried out in Chapter 2. We have examined the empirical distribution function of the number of jumps separately for several covariate groups and tried to identify the groups that are more affected by this problem. Based on these findings, we have proposed in Chapter 3 some general statistical models based on Poisson and Negative Binomial distributions.

Chapters 4 and 5, in contrast, are concerned with predictability of 'future' employment patterns from the 'present' situation in the labour market. The starting point in these two chapters has been an intuitively appealing individual specific assumption, viz., employment of an individual follows a Markov chain. From this assumption, we have derived the conditional distribution of the aggregate employment vector. The trace of the dispersion matrix of this distribution has been interpreted as a measure of employment mobility for a large class of transition matrices. Thus Chapters 4 and 5 together provide a new analytical tool for measuring employment fluctuation based on the aspect of predictability.

Chapter 6 examines the phenomenon of employment predictability from yet another angle. It focuses on the problem of predicting an aggregate regional employment rate. Traditionally, this aggregate employment rate is explained by regressing the rate on its determinants assuming constant variance of the disturbance term across observations. It is well known that the variance of the disturbance term in a regression is intimately related with the accuracy of the prediction of future values of the dependent variable. In this dissertation, we have tried to show that under some very simple intuitively appealing assumptions regarding individual employment behavior, linear regression model will not be an appropriate specification for the problem concerned. In fact, we have shown that in the present case, the dependent variable will be necessarily heteroscedastic, with its variance being a function of the explanatory variables. Since different subgroups of the population are likely to have different variance of the disturbance term, they will face different degrees of employment risk. Thus Chapter 6 provides yet another way of ordering employment risk of different subgroups in the population, The empirical part of this dissertation is concerned with the application of the above methods to the 38-th round of the NSSO data on employment and unemployment for rural Maharashtra. The methods considered yield broadly similar results, implying consistency of the empirical findings.

Broadly speaking, our empirical results suggest that so far as employment fluctuation is concerned, the persons from wage dependent households in rural Maharashtra are substantially more susceptible to employment fluctuation compared to the rest of the population. The two measures of employment fluctuation (e.g., average number of jumps and our proposed measure of predictability) are much higher for them than the rest of the population. Thus the wage dependent households in the rural sector change states more often and this change is more unpredictable than that for the rest of the population. The educated persons, in contrast, experience comparatively fewer state changes. Also, age and measures of employment fluctuation are found to be highly related. No significant relationship, however, is found between the measures of employment fluctuation and the other covariates in our study (e.g., region, social group etc.).

While these findings may be interesting, the limitations of the present study should be borne in mind. First, the empirical research reported here has been carried out for only one state of India, viz., Maharashtra mainly for illustrative purpose. To establish our findings more conclusively, further evidence from other states of India is required. Also, many other potentially important questions connected with the phenomenon of employment fluctuation have not been examined simply due to severe data constraints. For example, the 38-th round data on employment and unemployment collected by the NSSO, which have been used in this dissertation, come from a cross-section sample survey with a moving reference period. Although the problem of employment fluctuation can be addressed with these data, the span of the reference period is so short that they may not reflect the steady state employment positions of the households accurately.

We end this dissertation by pointing out some other potentially important research areas connected with this study. We have already mentioned these generalisations at the concluding section of the relevant chapters. Here we shall briefly recapitulate and summarise these possible extensions.

In Chapter 3 we have specified a Poisson regression model to explain the variation of number of 'jumps' of individuals. We have noted there the problems of overdispersion which is a common problem of such models. To tackle this problem of overdispersion a negative binomial model has been proposed and estimated. While the statistical fit of this latter model is found to be satisfactory, some discrepancies in the observed and the estimated values of the probabilities of one jump and two jumps have been noted. These discrepancies may occur if the *intensity parameter* λ changes with the number of jumps. If this be the case, the resulting process will no longer be a Poisson process and a closed form solution will be more difficult to obtain. However, such a process may also solve the problem of overdispersion.

Also, in Chapter 6 we have examined the consequences of aggregation of employment states over time and space. We have shown that under certain individual specific assumptions, the classical linear regression model for explaining variations in aggregate employment rate may not be valid. In that context, we have considered two alternative structures : (i) independence *and* (ii) Markov chains. We have merely examined the aggregation implications of these alternative structures and made no attempt to estimate these models from *individual specific data*. Such an exercise may help identify which of these alternative structures give better results.

These generalisations, once carried out, will undoubtedly throw new lights on the problem of employment fluctuation. However, we leave this as an agenda for future research.

References

- Abraham,A., 1980 : 'Maharashtra's Employment Guarantee Scheme'; *Economic and Political Weekly*, 15, 1339–1342.
- Akerlof,G.A. and Main,B.G.M., 1980 : 'Unemployment Spells and Unemployment Experience'; *American Economic Review*, 70, 885–893.
- Andrews,M.J., 1987 : 'The Aggregate Labour Market : An Empirical Investigation into Market Clearing for the UK'; *Economic Journal*, 97, 157–176.
- Ashenfelter,O. and Card,D., 1982 : 'Time Series Representation of Economic Variables and Alternative Models of the Labour Market'; *Review of Economic Studies*, 49, 761–782.
- Atkinson,A.B., 1970 : 'On the Measurement on Inequality'; *Journal of Economic Theory*, 2, 244–263.
- Atkinson,A.B., 1983 : 'The measurement of Economic Mobility'; in *"Social Justice and Public Policy"; M.I.T. Press, Cambridge.*
- Bagchee,S., 1984 : 'Employment Guarantee Scheme in Maharashtra'; *Economic and Political Weekly*, 15, 1975–1982.
- Bardhan,K., 1977 : 'Rural Employment, Wages and Labour Markets in India : A Survey of Research'; *Economic and Political Weekly*, 12, June 25, July 2 and July 9.
- Bardhan,P.K. 1978 : 'On Measuring Rural Unemployment'; *Journal of Development Studies*, 14, 342–352.
- Bardhan,P.K., 1982 : 'Little Girls and Death in India'; *Economic and Political Weekly*, 17, 1448–1450.
- Bardhan,P.K., 1984 : "Land, Labor and Rural Poverty"; *Columbia University Press.*
- Bardhan,P.K. and Srinivasan,T.N. [Ed.], 1974 : "Poverty and Income Distribution in India"; *Statistical Publishing Society.*
- Bardhan,P.K. and Srinivasan,T.N. [Ed.], 1988 : "Rural Poverty in South Asia"; *Oxford University Press.*

- Bartholomew, D.J., 1982 : "Stochastic Models for Social Processes"; *John Wiley and Sons, New York.*

- Behara,M., 1990 : "Additive and Nonadditive Measures of Entropy"; *Wiley Eastern Limited.*

- Behara,M. and Nath,P., 1973 : 'Additive and Nonadditive Entropies of Finite Measurable Partitions' in "Lecture Notes in Mathematics", *Springer-Verlag, Berlin*, **296**, 102–138.

- Bery,A. and Sabot,R.H., 1978 : 'Labor Market Performance in Developing Countries : a Survey'; *World Development*, **6**, 1199–1249.

- Bhattacharya,S., 1981 : 'Measurement of Employment and Unemployment in the Indian National Sample Survey A Review'; *Sankhya : The Indian Journal of Statistics*, **38**, Series C, 115–137.

- Binswanger,H., 1980 : 'Attitude toward Risk : Experimental Measurement in Rural India'; *American Journal of Agricultural Economics*, **62**, 395–407.

- Blanchard,O.J. and Mankiw,N.G., 1988 : 'Consumption : Beyond Certainty Equivalence'; *American Economic Review, Papers and Proceedings*, **78**, 173–177.

- Blau,D.M., 1987 : 'A Time-Series Analysis of Self-employment in the United States'; *Journal of Political Economy*, **95**, 445–467.

- Bliss,C.J. and Stern,N.H., 1982 : "Palanpur : The Economy of an Indian Village"; *Oxford University Press.*

- Blundell,R. and Walker,I. [Ed.], 1986 : "Unemployment, Search and Labor Supply"; *Cambridge University Press.*

- Bowers,N., 1980 : 'Probing the Issues of Unemployment Duration'; *Monthly Labour Review*, **103, July**, 23–32.

- Bronars,S.G. and Jansen,D.W., 1987 : 'The Geographic Distribution of Unemployment Rates in the U.S. : A Spatial-Time Series Analysis'; *Journal of Econometrics*, **36**, 251–279.

- Caballero,R.J., 1990 : 'Consumption and Precautionary Savings : Empirical Implications'; *Journal of Monetary Economics*, **25**, 113–136.

- Caballero,R.J., 1991 : 'Earnings Uncertainty and Aggregate Wealth Accumulation'; *American Economic Review*, **81**, 859–871.

- Cameron,A. and Trivedi,P., 1986 : 'Econometric Models Based on Count Data : Comparisons and Applications of Some Estimators and Tests'; *Journal of Applied Econometrics*, **1**, 29–53.

- Chakravarty,S.R., 1990 : "Ethical Social Index Numbers", *Springer-Verlag.*

- Chakravarty,S.R., 1994 : 'A Note on the Measurement of Mobility'; *Mimeo*, Forthcoming in *Economics Letters.*

- Conlisk,J., 1989 : 'Ranking Mobility Matrices'; *Economics Letters*, **29**, 231–255.

- Conlisk,J., 1990 : 'Monotone Mobility Matrices'; *Journal of Mathematical Sociology*, **15**, 173–191.

- Cooil,B., 1991 : 'Using Medical Malpractice Data to Predict the Frequency of Claims : A Study of Poisson Process Models with Random Effects'; *Journal of the American Statistical Association*, **86**, 285–295.

- Coondoo,D. and Dutta,B., 1992 : 'Measurement of Income Mobility : An Application to India', in "Development Policy and Economic Theory" by Basu,K. and Nayak,P. [Ed.], *Oxford University Press.*

- D'Agostino,R.B. and Stephens,M.A., 1986 : "Goodness-of-fit Techniques" ; *Marcel Dekker, Inc., New York and Basel.*

- Dandekar,V.M. and Rath,N., 1971 : 'Poverty in India'; *Economic and Political Weekly*, **6**, Nos. 1 and 2.

- Dantwala,M.L., 1979 : 'Rural Unemployment : Facts and Issues'; *Economic and Political Weekly*, **14**, 1048–1057.

- Dardanoni,V., 1993 : 'Measuring Social Mobility'; *Journal of Economic Theory*, **61**, 372-394.

- David,M. and Otsuki,T., 1968 : 'Forecasting Short-run Variation in Labor Market Activity'; *The Review of Economics and Statistics*, **50**, 68–77.

- Deaton,A., 1992: "Understanding Consumption" ; *Oxford University Press.*

- Denton,F.T., 1973 : 'A Simulation Model of Month-to-Month Labor Force Movement in Canada'; *International Economic Review*, **14**, 293–311.

- Dev,S.M., 1990 : 'Nonagricultural Employment in Rural India : Evidence at a Disaggregate Level'; *Economic and Political Weekly*, **25**, 1526–1536.

- Dev,S.M., 1992 : 'Poverty Alleviation Programmes : A Case Study of Maharashtra with Emphasis on Employment Guarantee Scheme', *Discussion Paper No. 77, Indira Gandhi Institute of Development Research.*

- Drèze,J.H. and Sen,A., 1990 : "Hunger and Public Action" ; *Clarendon Press, Oxford.*

- Dunn,R., 1987 : 'Analysing Spatial Time Series of Local Unemployment : A Geographical Approach Using Principal Component Analysis and Seasonal Adjustment Procedures'; *Environmental Planning*, A, **19**, 225–246.

- Dunsmuir,W. *et al.*, 1989 : 'Modelling of Transition between Employment States for Young Australians'; *Australian Journal of Statistics*, **31A**, 165–196.

- Dynarski,M. and Sheffrin,S.M., 1987 : 'Consumption and Unemployment'; *Quarterly Journal of Economics*, **102**, 411–428.

- Ezekiel,H. and Stuyt,J., 1990 : 'Maharashtra Employment Guarantee Scheme : Geographical Distribution of Employment' ; *Economic and Political Weekly*, **25**, A86–A92.

- Frydman,H., 1984 : 'Maximum Likelihood Estimation in the Mover-Stayer Model'; *Journal of the American Statistical Association*, **79**, 632–638.

- Fuller,W.A., 1975 : 'Regression Analysis for Sample Surveys'; *Sankhya : The Indian Journal of Statistics*, **37**, Series C, 117–132.

- Funke,M., 1992 : 'Time Series Forecasting of the German Unemployment Rate'; *Journal of Forecasting*, **11**, 111–125.

- Geweke,J., Marshall,R.C. and Zarkin,G.A., 1986 : 'Mobility Indices in a Continuous Time Markov Chain'; *Econometrica*, **54**, 1407–1423.

- Grogger,J., 1990 : 'The Deterrent Effect of Capital Punishment : An Analysis of Daily Homicide Counts'; *Journal of the American Statistical Association*, **85**, 295–303.

- Hasan,A. and de Broucker,P., 1982 : 'Duration and Concentration of Unemployment'; *Canadian Journal of Economics*, **15**, 735–756.

- Hausman,J. *et al.*, 1984 : 'Econometric Models for Count Data with an Application to the Patents R & D Relationship' ; *Econometrica*, **52**, 909–938.

- Hougard,P., 1987 : 'Modelling Multivariate Survival'; *Scandinavian Journal of Statistics*, **14**, 291–304.

- Judge *et al.*, 1985 : "The Theory and Practice of Econometrics"; *Wiley Series in Probability and Mathematical Statistics.*

- Just,R.E. and Pope,R.D., 1978 : 'Stochastic Representation of Production Functions and Econometric Implications'; *Journal of Econometrics*, 7, 67–86.

- Just,R.E. and Pope,R.D., 1979 : 'Production Function Estimation and Related Risk Considerations'; *American Journal of Agricultural Economics*, **61**, 276–284.

- Karlin,S. and Taylor,H.M., 1975 : "A First Course in Stochastic Processes"; *Academic Press, Second Edition, London.*

- Kemeny,J.G. and Snell,J.L., 1976 : "Finite Markov Chains"; *Springer, New York.*

- Klebaner,F.C., 1988 : 'The Unemployment Rate when the Labour Market is Modelled by a Markov Chain'; *Australlian Journal of Statistics*, **30**, 183–188.

- Konjin,H.S., 1962 : 'Regression Analysis in Sample Surveys'; *Journal of American Statistical Association*, **57**, 590–606.

- Krishna,R., 1973 : 'Unemployment in India'; *Economic and Political Weekly*, **8**, 475–484.

- Krishna,R., 1979 : 'Rural Unemployment : A Summary of Concepts' in "Agricultural Development of India" by Shah,C.H. and Vakil,C.N. [Ed.] *Orient Longman.*

- Kullback,S., 1959 : "Information Theory and Statistics", *John Wiley, New York.*

- Kynch,J. and Sen,A.K., 1983 : 'Indian Women : Well-being and Survival'; *Cambridge Journal of Economics*, **7**, 363–380.

- Lancaster,T., 1990 : "The Econometric Analysis of Transition Data"; *Cambridge University Press.*

- Lawless,J.F., 1987 : 'Regression Models for Poisson Process Data'; *Journal of the American Statistical Association*, **82**, 808–815.

- Lee,T.C., Judge,G.G. and Zellner,A., 1970 : "Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data", *North Holland, Amsterdam. London.*

- Leland,H.E., 1968 : 'Saving and Uncertainty : The Precautionary Demand for Saving'; *Quarterly Journal of Economics*, 8, 465–473.

- Maasoumi,E. and Zandvakili,S., 1986 : 'A Class of Generalized Measures of Mobility with Applications'; *Economics Letters*, 22, 97–102.

- Maasoumi,E. and Zandvakili,S., 1990 : 'Generalized Entropy Measures of Mobility for Different Sexes and Income Levels'; *Journal of Econometrics*, 43, 121–133.

- Maddala,G.S., 1983 : 'Limited Dependent and Qualitative Variables in Econometrics'; *Cambridge University Press.*

- Mahalanobis,P.C., 1968 : 'Concept and Measurement of Unemployment' in "Employment and Unemployment" by Rao,V.K.R.V. [Ed.]; *Allied Publishers.*

- Markandya,A., 1984 : 'The Welfare Measurement of Changes in Economic Mobility'; *Economica*, 51, 457–471.

- Minhas,B.S., 1988 : 'Validation of Large Scale Sample Survey Data : Case of NSS Estimates of Household Consumption Expenditure'; *Sankhya : The Indian Journal of Statistics*, 50, Series B, 1–63.

- Mullahy,J., 1986 : 'Specification and Testing of Some Modified Count Data Models'; *Journal of Econometrics*, 33, 341–365.

- Murthy,M.N. and Ray,A.S., 1975 : 'Development of the Sample Design of the Indian National Sample Survey Design During Its First Twentyfive Rounds'; *Sankhya : The Indian Journal of Statistics*, 37, Series C, 1–42.

- Nathan,G. and Holt, 1980 : 'The Effect of Survey Design on Regression Analysis'; *Journal of the Royal Statistical Association*, 42, Series B, 377–386.

- National Sample Survey Organisation, Government of India, 1988 : 'A Note on the Third Quinquennial Survey of Employment and Unemployment ; 38th Round (January – December 1983)', *Sarvekshana*, 11, 1–73.

- National Sample Survey Organisation, Government of India, 1988 : 'Report on the Third Quinquennial Survey of Employment and Unemployment, Maharashtra', No.341/4.

- Pal,P., 1989 : "Studies on Level of Living and Poverty in Rural India" ; *Ph. D. Dissertaion, Indian Statistical Institute.*

- Paul,S., 1988 : 'Unemployment and Underemployment in Rural India'; *Economic and Political Weekly*, **23**, 1475–1484.

- Paul,S., 1992 : 'An Illfare Approach to the Measurement of Unemployment'; *Applied Economics*, **24**, 739–743.

- Planning Commission, Government of India, 1970 : "Report of the Committee of Experts on Unemployment Estimates".

- Prais,S.J., 1955 : 'Measuring Social Mobility'; *Journal of the Royal Statistical Society*, **118**, Series A, 56–66.

- Rao,V.K.R.V., 1968 : "Employment and Unemployment", *Allied Publishers*.

- Ray,S.N. and Jacob,P., 1990 : 'Employment, Unemployment and Underemployment – Current Dimensions and Conceptual Issues : A Study Based on the NSS 43rd Round Results'; *IASSI Quarterly*, **9**, 111–128.

- Rosenzweig,M.R. and Evenson,R.E., 1977 : 'Fertility, Schooling and the Economic Contribution of Children in Rural India'; *Econometrica*, **45**, 1065–1079.

- Roumasset,J.A., Boussard,J.M. and Singh,I.J. [Ed.], 1978 : "Risk, Uncertainty and Agricultural Development'; *Agricultural Development Council, New York.*

- Salant,S.W., 1977 : 'Search Theory and Duration Data : A Theory of Sorts'; *Quarterly Journal of Economics*, **91**, 39–57.

- Sampson,M., 1990 : 'A Markov Chain Model for Unskilled Workers and the Highly Mobile'; *Journal of the American Statistical Association*, **85**, 177–180.

- Scott,A.J. and Holt,L.T., 1982 : 'The Effect of Two Stage Sampling on Ordinary Least Squares Methods'; *Journal of the American Statistical Association*, **77**, 848–854.

- Sen,A.K., 1966 : 'Peasants and Dualism with or without Surplus Labour'; *Journal of Political Economy*, **74**, 425–450.

- Sen,A.K., 1984 : 'Family and Food : Sex-bias in Poverty' in "Resources, Values and Development" by Sen,A.K., *Harvard University Press, Cambridge, Mass.*.

- Shorrocks,A.F., 1976 : 'Income Mobility and the Markov Assumption'; *Economic Journal*, **86**, 566–578.
- Shorrocks,A.F., 1978 : 'The Measurement of Mobility', *Econometrica*, **46**, 1013–1024.
- Shorrocks,A.F., 1992 : 'Spell Incidence, Spell Duration and the Measurement of Unemployment'; *Mimeo, University of Essex.*
- Shorrocks,A.F., 1993a : 'On the Hart Measure of Income Mobility' in "Industrial Concentration and Economic Inequality : Essays in Honour of Peter Hart" by Casson,M. and Creedy,J. [Ed.], *Edward Elgar.*
- Shorrocks,A.F., 1993b : 'On the Measurement of Unemployment'; *Mimeo, University of Essex.*
- Shukla,V., 1991 : 'Rural Non-farm Activity : A Regional Model and Its Application to Rural Maharashtra'; *Economic and Political Weekly*, **26**, 2587–2595.
- Sims,C., 1980 : 'Macroeconomics and Reality'; *Econometrica*, **81**, 1–48.
- Sommers,P.M. and Conlisk,J., 1978 : 'Eigenvalue Immobility Measures for Markov Chains'; *Journal of Mathematical Sociology*, **6**, 253–260.
- Sundaram,K. and Tendulkar,S.D., 1988 : "Toward an Explanation of Inter-regional Variations in Poverty and Unemployment in Rural India' in "Rural Poverty in South Asia" by Bardhan,P.K. and Srinivasan,T.N. [Ed.], *Oxford University Press.*
- Szal,R. and Robinson,S., 1977 : 'Measuring Income Inequality' in "Income Distribution and Growth in the Less Developed Countries" by Frank,C.R. and Webb,R.C. [Ed.], *Washington D.C. : The Brookings Institute.*
- Tano,D.K., 1991 : 'Are Unemployment and Out of Labour Force Behaviourally Distinct Labour Force States ? New Evidence from Gross Change Data'; *Economics Letters*, **36**, 113–117.
- Theil,H., 1972 : "Statistical Decomposition Analysis" ; *North Holland, Amsterdam.*
- United Nations, 1958 : "Handbook of Population Census Methods", Series F, No. 5.
- United Nations, 1984 : "Handbook of Household Surveys : Reviews and Methods", Series F.

- Unni,J., 1991 : 'Regional Variations in Non-Agricultural Employment : An Exploratory Analysis'; *Economic and Political Weekly*, **26**, 969–979.

- Vidwans,S.M., 1980 : 'A New Analysis of NSS Data on Rural Underemployment'; *Sankhya : The Indian Journal of Statistics*, **42**, Series B, 110–125.

- Visaria,P., 1981 : 'Poverty and Unemployment in India : An Analysis of Recent Evidence', *World Development*, **9(3)**.

- Visaria,P. and Minhas,B.S., 1991 : 'Evolving an Employment Policy for the 1990s : What do the Data Tell Us ?'; *Economic and Political Weekly*, **26**, 969–979.

- Walker,T.S. and Ryan,J.G., 1990 : "Village and Household Economies in India's Semi-arid Tropics"; *John Hopkins University Press, Baltimore, U.S.A.*

- Wedel,M. *et al.*, 1993 : 'A Latent Class Poisson Regression Model for Heterogeneous Count Data'; *Journal of Applied Econometrics*, **8**, 397–411.

- Zeldes,S.P., 1989 : 'Optimal Consumption with Stochastic Income'; *Quarterly Journal of Economics*, **104**, 275–289.

- Zellner,A., 1979 : 'Statistical Analysis of Econometric Models'; *Journal of the American Statistical Association*, 81, 628–643.

- Zellner,A. and Palm, 1974 : 'Time Series Analysis and Simultaneous Equation Econometric Models'; *Journal of Econometrics*, **2**, 17–54.

Appendix A

The Sampling Design of the 38-th Round Survey Operation of the NSSO

A.1 Introduction

The implications of ignoring the complexity of the survey design in forming estimates of parameters may turn out to be very crucial. This appendix, therefore, provides a short discussion on the sampling design adopted in the 38-th round survey operation of the NSSO.

Over the years, the survey design adopted by the NSSO has undergone many changes.[1] The survey design, even now, may change from round to round depending upon the nature of the enquiry. Even for the same round, the estimation procedure may be somewhat different for the rural and the urban sector.

A stratified two-stage sampling design was adopted in the 38-th round. For the rural sector, census villages were defined as the first stage units and households were defined as the second stage units. Sample villages were selected with *probability proportional to size and with replacement* (PP-SWR) in the form of two independent subsamples. In each of the selected first stage units, ten sample households were selected *circular systemati-*

[1]See Murthy and Ray (1975) for a brief history of the development of the designs adopted by the NSSO.

cally as second stage units after arranging all the households of a selected first stage unit in a specified manner.

Altogether the survey covered about 1.2 lakh households spread over more than 8600 sample villages and 4500 sample blocks as *central sample*. It covered the whole of India excepting (i) Ladakh and Kargil districts of Jammu and Kashmir *and* (ii) rural areas of Nagaland. The sample size for rural Maharashtra consisted of 568 villages. The number of households surveyed was 5388, with the total number of persons being 27756.

A.2 Survey Period and Subround Formation

The field work for the survey started in January, 1983 and was completed in December, 1983. The entire survey period of one year was divided into four subround periods of three months' duration coinciding approximately with the four agricultural seasons. The subround periods were, January to March 1983 (Subround 1), April to June 1983 (Subround 2), July to September 1983 (Subround 3) and October to December 1983 (Subround 4). The sample villages and blocks were distributed over the four subrounds in a manner so that valid estimates for each of the subrounds can be obtained separately.

A.3 Stratification and Allocation of Samples

All the states of India were first divided into some agroeconomic regions which were groups of contiguous districts, broadly similar with respect to population density and crop pattern.[2] Within each region the basic strata were formed so that they did not cut across districts. Each district with less than 1.8 million rural population, according to 1981 census, formed one basic stratum itself. A district with more than 1.8 million population was divided into two or more basic strata by grouping contiguous *tehsils* that are homogeneous, so far as possible, with respect to population density and crop pattern. The basic strata thus formed served as the ultimate strata for sampling in the rural areas.

[2] The number of regions considered for Maharashtra is six. Appendix C provides a short discussion on the definitions of these regions.

A total all-India sample of about 13170 first stage units was first allocated to each state (union territory) in proportion to the net investigator strength. This was further allocated to rural and urban sectors within each state (union territory) considering the relative size of its rural and urban population. All state allocations were rounded of to multiples of eight to have equal sample size in each of the four subrounds for either of the two subsamples into which the total population was divided. The rural (urban) allocations at state level were reallocated to strata in proportion to their rural (urban) population ensuring that the region level allocations were multiples of eight.

A.4 Estimation Procedure

The estimation procedure adopted in the NSSO 38-th round is based on a weighted sum of the sample values. Let θ be any population parameter of interest. Denoting by $\widehat{\theta}$ the unbiased estimate of θ (the state/region total of any variable), it is given by

$$\widehat{\theta} = \sum_{s} \frac{P_s}{n_s} \sum_{i} \frac{D_{si}}{p_{si} C_{si}} \frac{H_{si}}{h_{si}} \sum_{j} \theta_{sij} \qquad \text{(A.4.1)}$$

where

(i) the suffixes s, i and j stand for stratum, village and household respectively;

(ii) P_s and n_s give the stratum values of the selected population and number of sample villages respectively;

(iii) D_{si}, C_{si} and p_{si} stand for, number of hamlet groups formed, number of census villages contained in a larger revenue village actually surveyed and the selection population of the sample village in the order they are named;

(iv) H_{si} and h_{si} respectively denote the total number of households listed and the number of sample households;

(v) θ_{sij} is the observed value of the variate of a sample household.

The sample design described above is not *self-weighting.* Each sample household represents a different number of households in the population. Since these numbers, which are called *multipliers*, are different for different households, estimates of any aggregate should be based on a weighted sum of the samples, the weights being the multipliers.

Estimates of θ for a subsample (subround) can be obtained by restricting the summation ($\sum$) to units belonging to the concerned subsample (subround). Estimates of ratios such as percentages, averages etc. should be calculated by obtaining the unbiased estimates of the numerators and the denominators separately and then by division.

All estimates of parameters in this dissertation have been calculated by using the suitable multipliers as weights.

A.5 Maximum Likelihood Estimation

In the NSSO sampling design for the 38-th round, each sample household represents a different number of households in the population. Therefore, maximum likelihood estimation of any parameter of interest, say θ, should be carried out accordingly. We elucidate the basic principle with a simple example.

Suppose X and Y are two random variables. Let the joint probability density function (p.d.f.) of X and Y be $f(x, y, \theta)$. Let the weight function be $w(y)$ which is a function of y only. The weighted p.d.f. is given by,

$$f^{(w)}(x, y, \theta) = \frac{w(y)f(x, y, \theta)}{E[w(Y)]} \tag{A.5.1}$$

From the NSSO samples, we obtain observations on $(X^{(w)}, Y^{(w)})$ from the p.d.f. (A.5.1) and draw inference on θ.

Note that from (A.5.1) we can easily find out the marginal distribution ${f_X}^{(w)}(x, \theta)$ of $X^{(w)}$. Here,

$${f_X}^{(w)}(x, \theta) = \frac{w(x, \theta) f_X(x, \theta)}{E[w(X, \theta)]} \tag{A.5.2}$$

with the weight function

$$w(x, \theta) = \int f(y \mid x, \theta) w(y) dy \tag{A.5.3}$$

Suppose we have a sample of size n from (A.5.1)

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

Note that an unbiased estimate of $E(X)$ is given by,

$$\frac{E[w(Y)]}{n} \sum_{i=1}^{n} \frac{x_i}{w(y_i)} \tag{A.5.4}$$

This method of estimation was used in Chapter 3 and Chapter 4 of this dissertation.

A.6 Estimation of Standard Error

From Section A.4, it is obvious that the sampling design adopted by the NSSO is a complex one. Therefore, direct estimation of the standard errors of the parameters is difficult. However, for the purpose of easy estimation of the standard errors of the parameters, the sample was divided into some independent replicates of the same basic design.

Suppose the combined sample consists of r *replicates* or *interpenetrating subsamples*. Let the estimates of θ separately from the r replicates be given by $\widehat{\theta_1}, \widehat{\theta_2}, \ldots, \widehat{\theta_r}$. Then another estimate of θ combining all the subsamples is the simple average of all these replicates, i.e.,

$$\widehat{\theta} = \sum_{i=1}^{r} \widehat{\theta_i} / r \tag{A.6.1}$$

and $\text{Var}(\widehat{\theta})$, denoted by V_θ, can be estimated as

$$\widehat{V_\theta} = \frac{1}{r(r-1)} \sum_{i=1}^{r} (\widehat{\theta_i} - \widehat{\theta})^2 \tag{A.6.2}$$

The above method of estimation of the variance of an estimator has some obvious advantages. First, no special computer software is necessary. Second, it imposes no parametric or nonparametric estimation problem. Third, the cross validatory splitting of the data set into r parts is often recommended for data analysis.

The NSSO 38-th round *central sample*, consisted of two independent subsamples. Hence, here r reduces to two only. This means, the estimated standard error $s_{\widehat{\theta}}$ of an estimator $\widehat{\theta}$ becomes :

$$s_{\widehat{\theta}} = \frac{1}{2} \mid \widehat{\theta_1} - \widehat{\theta_2} \mid \tag{A.6.3}$$

Although this method provides an easy way to estimate standard error, it has only one degree of freedom and therefore may not be always reliable from the point of view of statistical inference. However, when the number of sample households in each subsample is large, the estimates are unlikely to differ much.

Let us define

$$t_o = \frac{\widehat{\theta}}{s_{\theta}} \qquad (A.6.4)$$

Note that here t_o follows t-distribution with only one degree of freedom. Hence, the tests of significance of the parameters are to be carried out accordingly.

Appendix B

NSSO Activity Codes

Table B.1.1: *NSSO Codes for Various Activities*

Description of the Activity Category	Category Code
(1)	(2)
Working with an employer under obligation but work not specifically compensated by any wage or salary	01-04
Worked (Self-employed) in household enterprise	11
Worked as a helper in household enterprise	21
Worked as a regular salaried or wage employee	31
Worked as casual wage labour in public works	41
Worked as casual wage labour in other type of works	51
Did not work though there was work in the family enterprise	61
Did not work but had regular salaried or wage employment	71
Sought work	81
Did not seek but was available for work	82

(Table B.1.1 continued in the next page)

(Table B.1.1 continued)

Description of the Activity Category	Category Code
(1)	(2)
Attended educational institutions	91
Attended domestic duties only	92
Attended domestic duties and was also engaged in free collection of goods, sewing, tailoring, weaving etc. for household use	93
Too young to work or to attend school or to seek employment	94
Old and disabled	95
Rentiers, pensioners, remittance recipients etc.	96
Beggars, prostitutes etc.	97
Others	98
Did not work due to temporary sickness (for casual labourers only)	99

The activity category codes 61, 71, 82 and 99 were used only for the current status approaches.

Persons assigned any of the activities listed under the category codes 01 to 71 were treated as 'working' (or employed'). Persons assigned activity category codes 81 or 82 were treated as 'seeking and/or available for work' (or unemployed). The remaining persons, i.e., those who were assigned any of the activity category codes 91 to 99 were treated as 'not available for work' (or not in the labour force).

The data on activity particulars of the persons were, however, tabulated only for population of age five and above.

Appendix C

Data Definitions

- **Break-up of Various Regions :** NSSO divides Maharashtra into six agroclimatic regions. These regions are formed with the hope that separate regionwise estimates of unemployment will provide a better and a clearer picture of the labour market situation in Maharashtra than any aggregate measure comprising of the whole State. The regions are,

  1. Coastal : The Coastal region consists of Greater Bombay and also the districts of Thane, Kulaba, Ratnagiri and Sindhudurg of Maharashtra.[1] This region is popularly known as the Konkan region of Maharashtra.
  2. Inland Western : The Inland Western region of Maharashtra consists of the districts of Ahmednagar, Pune, Satara, Sangli, Solapur and Kolhapur.
  3. Inland Northern : The Inland Northern part is formed with only three districts, viz., Nasik, Dhule and Jalgaon.
  4. Inland Central : This region of Maharashtra roughly coincides with the Marathwada region. It consists of the districts of Aurangabad, Parabhani, Beed, Nanded, Usmanabad and Jalna.
  5. Inland Eastern : This region of Maharashtra is formed with the districts of Nagpur, Wardha, Yavatmal, Amaravati, Buldana and Akola.

[1]Here we use the 1983 definitions of the districts. After 1983, some of the districts of Maharashtra were divided into smaller districts because of administrative reasons.

6. Eastern : This region is formed with only two districts, viz., Chandrapur and Bhandara.

   The Inland Eastern and The Eastern Region of Maharashtra form a part of what is more popularly known as the Vidarbha region of India.

- **Gainful Activity** : Gainful activity or work is the activity pursued by persons for pay, profit or family gain or in other words the activity which adds value to the national product. Normally, it is an activity which results in production of goods and services. However, the activities in 'agriculture' (i.e., all activities relating to industry division 0) in which the part or the whole of the agricultural production is used for own consumption and does not go for sale, are considered 'gainful'. Execution of household chores or social commitments etc., however, are not considered as 'gainful' activities. The activities such as prostitution, begging, etc. which may result in earnings are, by convention, not considered 'gainful'.

- **Workers (or Employed)** : Persons engaged in any gainful activity are considered as 'workers' (or employed). They are the persons pursuing any one or more of the eight situations listed under 00-71 under the list of activity categories.

- **Seeking or Being Available for Work (or Unemployed)** : Persons who, owing to lack of work, had not worked but either sought work through employment exchanges, intermediaries, friends or relatives or making applications to prospective employers or express their willingness or availability for work *under the prevailing conditions of work and remuneration*[2], are considered unemployed.

- **Labour Force** : Persons categorised as 'working' (or employed) and categorised as 'seeking or being available for work' (or unemployed) together constitute the labour force.

- **Out of Labour Force** : Persons categorised as neither 'working' nor 'seeking or being available for work' are considered to be engaged in

[2]My italics

non-gainful activities and categorised as 'out of labour force'. The persons falling under these category are students, those engaged in domestic chores, rentiers, pensioners, those living on alms, recipients of remittance, infirm and disabled persons, prostitutes and smugglers etc.

- **Worker (Self-employed) in Household Enterprise (Usual and Current Activity Code 11)** : Persons who are engaged in their own farm or non-farm enterprises are defined as self-employed, the term used to designate their activity status. There are different kinds of self-employed persons. Some may operate their enterprise without hiring any labour. Some others may normally work on their own but occasionally hire a few labourers. There is also a third category who, by and large, regularly run their enterprises by hiring labour. The first two groups of self-employed are called own account workers and the third, employers.

- **Worked as Helper in the Household Enterprise (Usual and Current Activity Code 21)** : The helpers are a category of working persons who keep themselves engaged in household enterprises, working full or part time and do not receive any cash payment in return for the work performed or any share of the family earnings from the enterprise. They are household members, a large number of them being related to the household head. They are dependent members of the household working for the family enterprise and getting food and shelter like other members of the household. One may also come across persons in a household who do not receive any cash remuneration for their work in the household enterprise but have a share of the family earnings out of such enterprise. Such persons are not 'helpers' but 'self-employed'. On the other hand, if any member of the household works in the household enterprise for which he is paid wage/salary, he is treated as an employee.

- **Worked as Regular Salaried/Wage Employees (Usual and Current Activity Code 31)** : *Persons working in others' farm or* non-farm enterprises (both household and non-household) and getting, in return, salary or wages on a regular basis (and not on the basis of daily or periodic renewal of work contract) are treated as regular salaried/wage employees. This category will include not only those getting time wage but also earners getting piece wage or salary and paid apprentices, both full time and part time.

- **Worked as Casual Wage Labour (Usual and Current Activity Code 41 and 51)** : Persons engaged in others' farm or non-farm enterprises (both household and non-household) and getting in return wages according to the terms of the daily or periodic work contract are treated as casual wage labour. In the rural areas and also in Government project sites, a type of casual labourers can be seen who normally engage themselves in public works taken up by the Government or the local bodies for construction of roads, digging of ponds etc. The casual labourers doing only such activities are considered casual labour in public works. The rest are casual wage labour in other types of work.

- **Had Work in Household Enterprises but Did Not Work (Current Activity Code 61) and Had Regular Salaried/Wage Employment but Did Not Work (Current Status Code 71)** : Persons engaged in household farm or non-farm enterprises or working in others' farm or non-farm enterprises as salaried/wage employees but absenting from work only temporarily due to either sickness or for enjoying leave or holiday or for other reasons belong to these two categories 61 and 71 respectively.

- **Attended Domestic Duties and Was Also Engaged in Free Collection of Goods, Sewing, Tailoring, Weaving etc. for Household Use (Usual and Current Activity Code 93)** : In the rural areas in general and in the tribal areas in particular, domestic work usually includes among others, a lot of work for free collection of vegetables, roots, firewood etc. and also spinning and weaving cloth for household use. Persons found to be spending time regularly in performing the above mentioned activities along with household chores, belong to this category.

- **Household** : A group of persons normally living together and taking food from a common kitchen form a household. In case of hostels, hotels, etc., each boarder with his dependents and guests (if any) was considered to constitute a separate household. The normally resident members include temporary stayaways but exclude temporary visitors or guests. If a person lives in one place and takes food from another, then he is considered to be a resident of the place where he lives.

- **Household Land Possessed :** The area of land possessed by a household on the date of interview is recorded in acres. It is defined as :

  Land Possessed = Land Owned+ Land Leased In− Land Leased Out

- **Household Type :** At the time of detailed enquiry for the 38-th round survey operation, each sample household was assigned a one digit 'household type' code. Out of the following five different household type codes, the one appropriate for sample household was chosen on the basis of *the main source of income of the household for the past 365 days prior to the date of the survey :*

Household self-employed in non-agricultural occupations	1
Agricultural labour households	2
Other labour households	3
Households self-employed in agricultural occupations	4
Other households	9

- **Household Industry/Occupation :** For each household there is a six digited code number of which the first three digits from the left refer to the appropriate principal 'Industry Group' and the next three digits refer to the relevant principal 'Occupation Family' of the Industrial and Occupational Classification respectively. For industry and occupation codes, the National Industrial Classification (NIC, Central Statistical Organisation, Government of India, 1962) and National Classification of Occupations (NCO, Central Statistical Organisation, Government of India, 1968) were used. Principal household occupation was that occupation which brought maximum earning to the household in the year preceding the date of enquiry. The industry in which the principal occupation was pursued was considered as the principal industry group of the household.

All households, however, could not be classified by industry and occupation. There were some non-worker households and households who failed to respond to the query about their occupation.

- **Household Group :** A one digited code number represents the group to which a particular household belonged, viz., Scheduled Tribes (1), Scheduled Castes (2), Neo-Buddhists (3) and others (9).
- **Household Religion :** Like household group, a one digited code number gives the religion followed by a household, viz., Hinduism (1), Islam (2) etc.

Appendix D

A Response to the Comments of the Referees

This appendix is added after receiving the comments and criticisms of the referees regarding the Dissertation. The referees have raised many interesting points and specific queries. In what follows, I shall try to answer those queries, attempt to clarify some of the problems and admit some of my errors.

The anonymity of the referees forces me to address them as Referee 1 and Referee 2 respectively. These names have, in fact, been given by the Ph.D. Committee of the Indian Statistical Institute. Since this appendix attempts to respond to the comments of both the referees, for any comment I have tried to indicate the name of the referee from whom the comment originated. Also for convenience, this response has been arranged topic-wise, rather than referee-wise.

• Lack of Economic Sophistication

Referee 1 has pointed out that the Dissertation lacks economic sophistication vis-a-vis statistical sophistication. In defense, I confess my helplessness vis-a-vis the huge scope of the topic concerned. As is pointed out by Referee 1, this is an "interesting and potentially important issue". And yet the fact is, very little work has so far been done *using Indian data*, especially the data collected by the National Sample Survey Organisation (NSSO). The nature of the data is such that they need some specific treatments and even a preliminary analysis brings out many interesting theoretical and empirical problems. Indeed, an analysis in such a case can not be complete and comprehensive in *one single* Ph.D. dissertation. Therefore, I tried to resolve the problem by narrowing down my focus to a more confined area.

My basic purpose was to measure employment fluctuation. This seemed to me to be the first step for a rigorous statistical and economic analysis. Among the desirable properties regarding measurement, I tried to concentrate on the aspect of prediction. Since prediction is basically a statistical problem, in the Dissertation I restricted myself on only the statistical aspect of the problem. In this context, I readily agree that any possible extension of economic theory consistent with the empirical observations in the Dissertation would have been a significant step forward. However, an investigation of this type along with the statistical aspects would have needed some huge collaborative work.

• Comments on the Assumption of Markov Chain

Among the general observations, Referee 1 has questioned the justification of the assumption of Markov chain regarding day-to-day changes in the employment status of an individual. In this context, Referee 1 has observed that even a sketch of a simple behavioral model of individual labor supply along with some mechanism of job assignment would have been of immense help.

The justification of the assumption of Markov chain invoking economic theory is not an easy task. Generally, the specification of a Markov chain is made on an ad-hoc basis. It is based on two assumptions the validity of which is neither theoretically nor empirically evident : (1) the transition probabilities are constant over time (i.e., the Markov chain is stationary); (2) the probability of moving from one state to another is independent of past history (i.e., the Markov chain is of first-order). One can of course relax these assumptions. However, any such relaxation leads to quite a complicated stochastic process where the tractability of the standard model may be unrecoverably lost.

Often, such assumptions are made because the nature of the data-set under consideration is such that it lends itself amenable for such specifications. Consider a stochastic process $\{X_t\}$. If X_t's can take only a few values and if independence over time turns out to be a bad assumption, the assumption of Markov chain may be considered as a starting exploration. The first-order stationary Markov chain model provides an analytically tractable framework for a fruitful analysis of mobility. However, the 'price' for the nice structure is the two strong assumptions noted above. Therefore, the decision regarding whether or not a Markovian approach should be adopted for a particular problem is a subjective and context-dependent decision.

In my case, it was motivated from the corresponding literature on income fluctuation. Markov chain models have been used widely for addressing the problem of income fluctuation. Please note that even for income mobility, the specification of a Markovian model is not without problems (Fields and Ok, 1996).

The Markovian assumption made in the dissertation is more appropriate for the persons who do not have long time employment contracts. Thus, the model may be a good approximation for casual laborers. It can also be a good approximation for the housewives whose decision to join the labor force on a particular day is independent of the distant past. On the other hand, for salaried persons it may not reflect reality accurately. However, even if employment of salaried persons depends on distant past, the estimated transition probabilities will still be a crude approximation of the true situation.

• Comments on Chapter 2

Is Stability the Overwhelming Pattern ?

Regarding Chapter 2, Referee 1 has observed that "stability rather than fluctuation" seems to be the overwhelming pattern. This, however, is not true for all population groups. Referee 1 himself has observed that the lowest frequency of zero jumps in all the tables in Chapter 2 is 0.7851 for agricultural laborers in Table 2.4.4. Given the fact that these are day-to-day fluctuations, not month-to-month or year-to-year fluctuations, the probability seems to be quite low. On the other hand, the employment pattern for many other covariate groups, especially that for the self-employed, does indeed reflect stability.

Trends in Employment Vector : Some Possible Explanations

As reported in Table 2.3.1 of the Dissertation, the proportion of employed persons for the whole population changed from 0.4572 to 0.4668 within a span of seven days. The fluctuation of aggregate unemployment rate is even more prominent : unemployment rate decreased from 0.0730 to 0.0611 within a span of seven days.

Later, Referee 1 has demanded a possible explanation of the observed monotonic increase in employment from day 1 to day 6 and asked whether this could be due to some feature of the canvassing of the schedules. Following the advice of Referee 1, I again examined the schedule of enquiry that was canvassed. However, to the best of my ability, I have not been able to trace a feature in the schedule which may readily explain such pattern.

Please note that although aggregate employment increases from day 1 to day 6, it undergoes a fall in day 7. Moreover, the patterns over subrounds are also heterogeneous, as can be seen in Figure 4.5.1 (Chapter 4, pp. 74) of the Dissertation. To be specific, the trend in aggregate employment is absent in Subround 1, it is moderate in Subrounds 2 and 3 and is prominent in Subround 4. If there were any systematic error in measurement because of some features of the schedule, possibly it would have got reflected somewhat more uniformly across the subrounds.

Of course, this does not rule out the possibility of presence of some other kinds of non-sample biases,[1] like 'recall error' of individuals. *The sample individuals covered in the survey were not actually observed for seven days, only their employment experience for the past seven days were recorded as told by them.* Even if recall error does take place, there is no a priori reason why it should lead to an almost monotonic increase in employment and not in the states of unemployment and out of labor force !

However, there are two possible statistical explanations. First, suppose there exists a steady state level of employment, unemployment and out of labor force for the population and aggregate employment vector fluctuates around that steady state level. A random shock will occasionally push the employment vector away from the steady state and it will take some time to come back to the steady state. Possibly, the time required will be more than a week.[2] An approximate idea about the time required may be formed from the transition probabilities reported in Chapter 4.[3]

We may calculate *the asymptotic half-life to convergence to steady state* using the measure proposed by Shorrocks(1978) :

$$h = -\frac{\log 2}{\log \mid \lambda_2 \mid} \tag{D.0.1}$$

The interpretation of the above measure is as follows : consider the 'present' probability vector and the steady state probability vector. Suppose the distance between these two vectors is measured using a linear metric. Then, on an average, after h periods the distance between the two vectors will be half the present distance.

[1] Please note that the main purpose of my examination of the day-to-day movement of aggregate employment was precisely to check the internal consistency of the data.

[2] Note that the speed of convergence to steady state will be different for different population groups. Since the probability of jump for the laborers is high, it will converge to steady state quickly. However, for some groups like self-employed, the adjustment will be slow. A detailed examination regarding the speed of adjustment can possibly be a part of future research agenda.

[3] Please note that these transition probabilities have been used for the purpose of illustration only. They have been calculated for the entire population without distinguishing the effects of several covariate groups. Thus the idea on speed of convergence gleaned from them can only be approximate.

The values of h for the four subrounds are calculated as 11.26, 9.93, 8.10 and 9.32 days respectively. Please note that the gaps between the initial probability and the steady state probability of employment for the four subrounds are 0.0091, 0.0294, 0.0154 and 0.0503 respectively (Table 4.5.1, Chapter 4, pp. 72). Thus, convergence to steady state may take a longer time, especially for Subround 4.

This means if we observe the aggregate employment vector for a short span, we may observe trend in the data. This trend may occur because of the slow speed of convergence to the steady state. The observed trend may reflect only a part of the cycle around the steady state.

Second, aggregate employment may itself be non-stationary. In fact, this possibility has been mentioned in Chapter 4 of my Dissertation. There we have argued that even if individual employments are stationary, aggregate employment may not be so. However, the non-stationarity discussed in Chapter 4 comes from the variance, not from the mean.

Tests of Equality of Empirical Cumulative Distribution Functions

Referee 1 pointed out that the tests of significance of the differences among groups in each of the Tables 2.4.2 to 2.4.7 were not carried out. This was not done for the following reason : it was thought that the descriptive data analysis reported in Chapter 2 would help to motivate the readers gradually towards the central problem of the Dissertation. Therefore, I tried to keep Chapter 2 very simple and presented only some descriptive measures based on moments in it. All the statistical tests and models were presented in the later chapters.

The non-parametric tests for examining the equality of two distribution functions from their empirical counterpart are easy to implement in Chapter 2. However, in the analysis reported in Chapter 2, we considered only one covariate at a time. This was unlikely to yield the true partial effect of each covariate in the presence of significant interaction effect. In Chapter 3, such interactions were examined and the tests with respect to the parameters were carried out there.

Following the advice of Referee 1, we have carried out some non-parametric tests. To be specific, we have used Kolmogorov-Smironov (KS) two-sample tests. The KS test, based on the empirical cumulative distribution functions (CDF) of two different sets of samples, examines whether two independent samples have been drawn from the same population. The tests carried out here are all two-tailed tests. These tests are sensitive to any kind of difference in the distributions from which the two samples are drawn. If the two samples have been drawn from the same population distribution, the cumulative distributions of both samples may be expected to be close to each other. If, on the other hand, the two sample cumulative distributions are 'too far apart' at any point, this suggests that the samples come from different populations.

The two-sample KS test is based on the *maximum* of the observed differences of the two empirical CDF's. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be two sets of independent samples with empirical CDF F_m and F_n respectively. The two-sample KS test uses the statistic

$$D = \text{Maximum } \mid F_m(X) - F_n(Y) \mid \tag{D.0.2}$$

for a two-tailed test. The sampling distribution of D is known. The probabilities associated with the occurrence of values as large as an observed D under the null hypothesis that the two samples have come from the same population can be calculated. When both m and n are greater than 40, a large sample approximation of a KS test may be used. Critical values of D for the large sample two-tailed test at 5% and 1% level of significance are $1.36\sqrt{\frac{m+n}{mn}}$ and $1.63\sqrt{\frac{m+n}{mn}}$. Note that the critical values depend on the values of m and n. Some of these tests corresponding to the empirical CDF's reported in Chapter 2 of the Dissertation have been presented in Table D.1.1 of this response. The tests are broadly in agreement with the main conclusions of the Dissertation.

The tests do not indicate that the empirical CDF's are significantly *different* across subrounds. Among six such tests, empirical CDF of jump for Subround 1 turns out to be different from that of Subround 3 and Subround 4 at 5% level of significance. However, at the level of 1% they are found to be similar.

Among occupational categories, the KS tests reveal that the empirical CDF's corresponding to Selfagr, Selfnagr and Othocc are not significantly different. These three categories are, however, found to be different from Agrlab and Othlab at 1% level. The category LAB itself is not homogeneous – the empirical CDF's corresponding to Agrlab and Othlab are found to be different at 1% level.

Among Social Groups, SC and ST are found to form homogeneous groups. It is clearly different from the categories Othhind and Othmusm. Othhind and Othmusm are homogeneous among themselves. However, the category Othgrp is very similar to SC and ST. This is because a significant portion of this other group consists of Neo-Buddhists who are mostly converted SC's and ST's.

Across sex, the categories are broadly similar. Note that this is consistent with the findings from the Poisson regression model. However, agewise differences are more pronounced. Also, the empirical CDF's are widely heterogeneous across educational categories. All the three categories, Illit, Primid and Highed differ pairwise at 1% level of significance.

Comments on Choice Based Classifications

Referee 1 has pointed out that "to the extent occupation, whether or not to be a wage worker, even the region of residence are matters of choice by individuals, some of the groups are choice based." This, however, is true only in the long run. The data analyzed in the Dissertation was only for seven days and it is extremely unlikely that these choices, if at all made, were made during those seven days ! However, I feel that I should have stated that assumption explicitly and I thank Referee 1 for raising this interesting point.

Comments of Referee 2 on Chapter 2

On Chapter 2, Referee 2 has mentioned two possible sources of generalizations. Among them, the first relates to whether the seven days employment fluctuations used in this analysis is representative of employment fluctuations *over a longer period*. A detailed study of week-to-week or even month-to-month employment fluctuation will obviously be very interesting. However, there are several impediments, the main one being lack of such information in India.

The NSSO surveys on employment and unemployment are basically cross-section surveys. Each household, included in the survey is visited only once. Thus, though NSSO collects some information on employment of individuals pertaining to one year (i.e., measures based on usual status), *these information are available for one period only.*

Since measuring unemployment can be extremely complicated in an agro-based developing economy, it was felt that only the usual status measures would not reveal all the aspects of rural labor market. Therefore, some measures based on 'current status' was developed. That is why, side by side with the yearly experience, employment experience for the current week is also recorded by the NSSO. The weekly status of an individual is determined from these observations. Like usual status, this information is also available for only one period. Therefore, a panel data consisting of week to week or month to month fluctuation in employment cannot be obtained from the NSSO surveys directly.

However, one broad compensation is the information on the time spent *within the reference week.* In order to accurately measure the time spent on different activities, the employment experience of the individuals are recorded for each day of the reference week. These are the data that we have used in the Dissertation. In the existing set up, the best one can do is to calculate the day-to-day transition probabilities and project them to find out the corresponding week-to-week or month-to-month transition probabilities.

Regarding the day-to-day employment fluctuations, Referee 2 has indicated that even within a week "certain day to day fluctuations are more important than others" and that he would like to know whether each day of the week is treated similarly by individuals. Given the nature of the data, the answer would be extremely difficult to find out.

For collection of information on employment and unemployment, NSSO adopts a *moving* reference point. The households are not visited simultaneously. Moreover, the calender date on which data on a household is collected, although available in the questionnaire, is not punched to store in computer. Hence, one source of valuable information is lost.

The second possible source of generalization in Chapter 2 is an examination of individual risk vis-a-vis household risk. This is a stimulating topic which needs further investigation. Employment of one member of a household can possibly hedge the risk of unemployment of other members. In this context, the assumption of independence of employment of individuals is a crude assumption. However, it simplifies the model specification considerably.

Employment fluctuation and employment risk are, however, not exactly interchangeable terms, although they are closely related. Employment fluctuation is an individual-specific concept. It includes movements from or to the state 'out of labor force'. On the other hand, when we discuss employment risk, we restrict our attention to those who are in the labor force. Also, when one relates employment risk to the intertemporal consumption decisions, households and not individuals form the basic units. However, to assess the employment risk of an household correctly, the nature and extent of fluctuation of employment of all the members of the household is to be determined accurately.

A detailed examination of employment fluctuation of individual members of the households is, thus, the first step. That is why, throughout the major part of the Dissertation, we discussed employment fluctuation and only in Chapter 6, the aspect of employment risk was briefly touched upon.

To examine the nature and the extent of employment risk of households under this framework, several problems are to be solved. One such problem is unequal number of members across households. For example, to extend the analysis done in Chapters 2 and 3 of the Dissertation for households, one may calculate average number of jumps across households. However, average jump is not an integer variable. Therefore, a simple ordinary least squares (OLS) model for average jump on different covariate categories will perhaps be enough.

More general models of employment fluctuation across households should ideally take into account the correlation structure of employment states among the members of the household. Once again, specification of this structure is extremely difficult because of unequal number of members across households. Next, one should propose an aggregate measure of employment fluctuation for households and examine its properties. Given that some of the covariates like age, sex and education are individual specific, construction of such aggregates will also be extremely difficult. However, once such a task is accomplished, one may examine its relationship vis-a-vis the consumption decision of a household. For such an examination, ideally one needs panel data on consumption. NSSO surveys are cross-section surveys. Although NSSO collects the data on consumption and employment for the *same set of households*, such data are collected for one period only. Thus, the extent of consumption smoothing cannot be examined using NSSO data – unless one makes some dangerous assumptions.

In this context, Referee 2 has mentioned the work of Townsend(1994). Townsend did some fundamental work on consumption smoothing in village India. He tested the full insurance model using the *panel data* collected by the International Crop Research Institute for Semi-Arid Tropics (ICRISAT). I am not familiar with this data set and I do not know whether information on employment, income etc. are separately available across individuals. Therefore, I am not very sure whether same methods of analysis can be applied to the ICRISAT data set.

• Comments on Chapter 3

The Logit Model for Jump

Referee 1 has observed that "Given that the probability of zero jump is high, it would have been interesting to estimate a very simple probit of the probability of no jump".

It is true that here the data distinguish jump states from no jump states much more sharply than they distinguish the number of jumps between 1 and 6. However, this is only because of the short span of the reference period. The number of jumps committed by an individual would not at least decrease had the reference period been longer. Hence, the frequency of zero and one jumps could only decrease if the reference period was longer. As a consequence, the probability of two or more jumps would also increase at the cost of zero or one jump. In such a case, a probit or a logit model would not be a good approximation. However, the Poisson regression models are sufficiently flexible as they can take into account the specific nature of this type of data.

Moreover, although the data distinguish jump states from no jump states sharply, *the extent of sharpness is different for different population groups.* For example, the probability of two or more jumps is 0.1134 for agricultural laborers and 0.1043 for other laborers (Table 2.4.4, Page 38) – much higher than 0.0653, the probability for the general population.

A probit model which summarizes the variable 'number of jump' to an indicator variable would not have taken into account the differences among more than one jump across covariate groups. A Poisson regression model and its subsequent generalizations which take into account this difference seems to be a more appropriate model. Moreover, please note that among the different models estimated the probability of zero jump is estimated almost accurately by the negative binomial model, the highest value of PPE_0 for this model being only 0.002 (Table 3.3.3, Page 58) for Subround 1.

However, as suggested by Referee 1, we have estimated the logit model and the results are presented in Table D.1.2. An empirical comparison of the performance of a logit model with the Poisson regression model reveals some interesting features. The logit model for 'jump' versus 'no jump' was separately estimated for the subrounds. The same explanatory variables as in Poisson regression models were used. An interesting feature of the result is that the coefficients obtained from the logit model are broadly similar to that of Poisson regression models. For some variables, like LAB, the estimated coefficients are very close. The signs of the coefficients are also in agreement. However, the model does not fit as well as the negative binomial model. Please note that the observed relative frequency of zero jump in Subround 1, Subround 2, Subround 3 and Subround 4 are 0.8934, 0.8827, 0.8649 and 0.8684 respectively (Table 2.4.3, Chapter 2, pp. 37 of the Dissertation). The estimated relative frequencies of zero jump from the logit models for those four subrounds are 0.9688, 0.9515, 0.9592 and 0.9638 respectively. Thus, while the Poisson model underestimates the probability of no jump (Figure 3.3.1, Chapter 3, pp. 56), the logit model grossly overestimates it. The value of the measure PPE_0, as defined in Chapter 3 of the Dissertation, is higher for all the four subrounds.

Comments on Estimates of Standard Error

Referee 1 has questioned the reliability of the estimates of the standard errors calculated from the subsamples and suggested the use of large sample variance-covariance matrix for the parameter estimates based on Cramer-Rao information limit. Empirical comparisons of these two different types of estimates have been done by Coondoo(1975). The results are, however, mixed and do not indicate any pattern.

Although estimates of standard errors based on only two independent and interpenetrating subsamples may be somewhat inadequate, they have some definite advantages. First, it provides a direct check on the robustness of the results which cannot be obtained by other methods. Second, the formulae for subsample based standard errors do not depend on the design of the survey. If the sample design is complicated, formulae for the estimates of standard error are also bound to be complicated and may change if the design of the survey is changed. *The established computer packages that deal with different types of statistical problems generally do not take this aspect into consideration.* For empirical analyses based on survey data, researchers often use these computer packages and are forced to accept estimates of standard errors based on formulae which ignore the sample design.

The method based on subsamples, on the other hand, provides a very easy way of computing estimates of standard error. If the sample size in all the subsamples are large, the resulting estimates may turn out to be reasonably accurate.

Comments on Use of Sample Weights

Regarding my use of sample weights (multipliers) in estimation, Referee 1 has expressed his doubt by stating that it is not obvious whether weights should be used in this context. Even in the context of ordinary regression, the issue of reweighting is highly debated (Deaton, 1994). The decision to use multipliers in this context thus seems to have remained mostly a matter of subjective choice.

Comments of Referee 2 on Chapter 3

On Chapter 3, Referee 2 has observed several possibilities. I am thankful to him for his comments, especially for the reference of Gourieroux *et al* (1984a & b). Gourieroux *et al* has proposed the theory of pseudo maximum likelihood and applied it to a large class of models. The insights from their research may indeed be useful for further developments of the Poisson regression model proposed in the Dissertation. In this context, Referee 2 has also raised some queries regarding the possible incorporation of the boundedness of the observed number of counts (jumps). The Poisson and the negative binomial models, theoretically, can take values upto infinity. Therefore, the expected probabilities did not sum to one for Negative Binomial model as there exists some positive probability for more than six jumps. However, the estimated tail probabilities beyond six jumps are so small that the models may be considered as a fair approximation (See Figure 3.3.1, Chapter 3, pp. 56 of the Dissertation). Therefore, improvement in terms of goodness of fit for the truncated versions of Poisson and Negative Binomial models will only be very marginal.

• Comments on Chapter 4

Referee 1 has expressed his doubt regarding the assertion of independence of V_{it}^{*} given $\widetilde{X_{it}}$. A proof is presented below.

To avoid complex notations, we shall consider the case $K = 2$. The general proof for any K is similar and can be carried out in an exactly similar manner.

Let X_{it} be a random variable which denotes the employment position of the i-th individual on the t-th period ($i = 1, 2, \ldots, n$; $t = 1, 2, \ldots, T$). For convenience, we shall call each such period a *day*. We define,

$$\begin{aligned} X_{it} &= 1 \qquad \text{if the } i\text{-th individual is employed on } t\text{-th day} \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{D.0.3}$$

Let $\widetilde{X_i}$ be the employment vector corresponding to the i-th individual. Then,

$$\widetilde{X_i}' = (X_{i1}, X_{i2}, \ldots, X_{iT}) \qquad i = 1, 2, \ldots, n \tag{D.0.4}$$

Let us denote

$$\tilde{Y} = \sum_{i=1}^{n} \widetilde{X_i} \tag{D.0.5}$$

and

$$\tilde{Z} = \frac{1}{n} \sum_{i=1}^{n} \widetilde{X_i} \tag{D.0.6}$$

Here, $\tilde{Y}$ and $\tilde{Z}$ represent the vector of aggregate and proportion of employed persons on each day. Let us denote $Y_t = \sum_{i=1}^{n} X_{it}$ and $Z_t = \frac{1}{n} \sum X_{it}$. Here, $(Y_1, Y_2, \ldots, Y_T)$ or $(Z_1, Z_2, \ldots, Z_T)$ gives us a time series.

Let the initial probability vector be

$$\tilde{\pi}_0{}' = (\pi_0, 1 - \pi_0) \tag{D.0.7}$$

where π_0 is the initial probability of employment.

The transition matrix is given by

$$\mathrm{P} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} \tag{D.0.8}$$

where $p = P[X_{t+1} = 1 \mid X_t = 1]$ and $q = P[X_{t+1} = 0 \mid X_t = 0]$.

Also, let us assume that a unique steady-state distribution exists and is given by

$$\tilde{\pi}^{*\,\prime} = (\pi^*, 1 - \pi^*) \tag{D.0.9}$$

Let π_t be the probability that the i-th individual will be employed on the t-th day. Then it can be easily proved that

$$\pi_t = P[X_{it} = 1] = \pi^* + (p + q - 1)^t (\pi_0 - \pi^*) \tag{D.0.10}$$

Let us define a set of random variables V_{jk}, $(j = E, N; k = E, N)$, V_{jk} being the total number of persons who are in state j on day t and state k on day $(t+1)$. Then,

$$V_{EE} + V_{EN} + V_{NE} + V_{NN} = n \tag{D.0.11}$$

and

$$\begin{aligned} Y_t &= V_{EE} + V_{EN} \\ Y_{t+1} &= V_{EE} + V_{NE} \end{aligned} \tag{D.0.12}$$

Let

$$\tilde{V} = (V_{EE}, V_{EN}, V_{NE}, V_{NN}) \tag{D.0.13}$$

Then

$$\tilde{V} \sim \text{ Multinomial}\,(n, \pi_t p, \pi_t(1-p), (1-\pi_t)(1-q), (1-\pi_t)q) \tag{D.0.14}$$

Clearly

$$\begin{aligned} Y_t &\sim \text{ Binomial}\,(n, \pi_t) \\ Y_{t+1} &\sim \text{ Binomial}\,(n, \pi_{t+1}) \end{aligned} \tag{D.0.15}$$

Proposition D.0.1 *Let $g_1(V_{EE} \mid Y_t)$ and $g_2(V_{NE} \mid Y_t)$ are the conditional distribution of V_{EE} and V_{NE} given Y_t respectively. Also, let $g(V_{EE}, V_{NE} \mid Y_t)$ be the joint distribution of V_{EE} and V_{NE} given Y_t. Then,*

$$g(V_{EE}, V_{NE} \mid Y_t) = g_1(V_{EE} \mid Y_t) \cdot g_2(V_{NE} \mid Y_t) \tag{D.0.16}$$

i.e., given Y_t, V_{EE} and V_{NE} are independent.

Proof : Let the joint probability mass function (PMF) of $\widetilde{V}$ be $h(\widetilde{V})$. Since, $Y_t \sim$ Binomial (n, π_t) therefore if the PMF of Y_t be denoted by $\psi(Y_t)$ then

$$\psi(Y_t) = {}^{n}C_{Y_t}(\pi_t)^{Y_t}(1-\pi_t)^{n-Y_t} \tag{D.0.17}$$

This means,

$$\begin{aligned} g(V_{EE}, V_{NE} \mid Y_t) &= \frac{h(V_{EE}, Y_t - V_{EE}, V_{NE}, n - Y_t - V_{NE})}{\psi(Y_t)} \\ &= {}^{Y_t}C_{V_{EE}}(p)^{V_{EE}}(1-p)^{Y_t - V_{EE}} \times \\ &\quad {}^{n-Y_t}C_{V_{NE}}(1-q)^{V_{NE}}.(q)^{n-Y_t-V_{NE}} \end{aligned} \tag{D.0.18}$$

Now note that

$$g_1(V_{EE} \mid Y_t) = {}^{Y_t}C_{V_{EE}}(p)^{V_{EE}}(1-p)^{Y_t - V_{EE}} \tag{D.0.19}$$

and

$$g_2(V_{NE} \mid Y_t) = {}^{n-Y_t}C_{V_{NE}}(1-q)^{V_{NE}}.q^{n-Y_t-V_{NE}} \tag{D.0.20}$$

Therefore,

$$g(V_{EE}, V_{NE} \mid Y_t) = g_1(V_{EE} \mid Y_t) \cdot g_2(V_{NE} \mid Y_t) \tag{D.0.21}$$

i.e, given Y_t, V_{EE} and V_{NE} are independent. ■

Note that V_{EE} and V_{NE} are not independent. This is because, if the value of Y_t is not known, then V_{EE} and V_{NE} both can take values between 0 and n (depending upon the value of Y_t). However, if the value of Y_t is not known, V_{EE} and V_{NE} will be negatively associated because value of one variable bounds the highest possible value of the other. In fact,

$$\text{Cov}\,(V_{EE}, V_{NE}) = -n\pi_t(1-\pi_t)p(1-q) < 0 \tag{D.0.22}$$

• Comments on Chapter 5

Lack of Economic Motivation in the Proposed Mobility Index

Referee 1 has pointed out the lack of economic motivation or interpretation in the proposed mobility index. Mobility indices have been widely used to address the problem of income-fluctuation of individuals. Generally, such studies start with the specification of a Markov chain. Income is classified into some groups. Each such group forms a *state* of the corresponding Markov chain. Accordingly, some studies start with the specification of an intertemporal utility function and examine the welfare implications of the different mobility structures (Dardanoni, 1993). These studies may not even propose an index for mobility. However, they specify some well defined criteria for *ordering* the transition matrices.

Please note that a major distinction of the present study from those on income mobility is that in the latter case, *the states are ordered with respect to income.* However, in the case of employment mobility, a real difficulty would occur if one tried to order 'unemployment' vis-a-vis 'out of labor force'. No universal ordering seems to be possible because the choice will depend on individuals' reservation wage. This is one reason why the mobility indicators which do not focus on welfare implications offer different types of ordering from those focussing on the aspect of welfare (Section 5 of Dardanoni, 1993). The income mobility literature is still far from being unified on how to measure mobility. Some economists believe that "this is due to the wide gap (with a few notable exceptions) between those who devise measures of income mobility and those who measure mobility empirically" (Fields and Ok, 1996). Since measures of welfare were not directly relevant to our study, we ignored these aspects altogether.

Even if one ignores the aspect of welfare, the concept of mobility ordering is ill posed unless one systematically postulates properties which a reasonable index of mobility should satisfy. Shorrocks(1978) considered this approach. Geweke *et al.* (1986) have shown that Shorrock's criteria fall into logically distinct categories. Criteria within each category are mutually consistent. However, there are several conflicts across categories. Geweke *et al.* tried to remove these inconsistencies in a continuous time framework. The moot point from Geweke *et al.* and from Table 5.2.1 (Chapter 5, pp. 84) of the Dissertation (taken from Dardanoni (1993)) is : no single mobility statistics has the minimum requirements regarded as essential. Accordingly, appropriate mobility indices should be used according to the "need" of the specific context.

Our study focussed on the aspect of prediction of aggregate employment position in the 'next' period. Since prediction is basically a statistical problem, it can be given some simple statistical interpretations. However, it will be extremely difficult to offer any 'economic' interpretation. I, therefore, can only express my helplessness.

Lack of Test of Significance

Referee 1 has indicated that "average mobility indices of different sub-groups are compared but without any tests of the significance of the differences". Please note that I have not only presented the average mobility for the combined sample, but also separately for the subsamples. The standard errors (and hence the t-ratio's) can be calculated using the same method adopted in the Dissertation. Following the advice of Referee 1, these t-ratio's have been calculated and presented in Table D.1.3. The t-ratio's display considerable variation. This somewhat weakens the results obtained.

• Comments on Chapter 6

Regarding the motivation for Chapter 6, I feel that some clarifications are needed. It is not that I feel that the employment status variables X_{ijt} are independent *in reality*.[4] My purpose of investigation was : *even if we assume such a simplifying assumption, can the ordinary least squares (OLS) of aggregate employment rate on other covariates be justified ?* The answer obtained unambiguously stated that it is not.

In this context, Referee 1 has pointed out several possible generalizations. If we relax the assumption of independence across individuals, a correlation structure across individuals need to be specified. The basic unit of investigation then will be a household (village) and a possible generalization will be employment across households (villages) are independent.

Suppose, we assume that employment status of individuals in household i is independent of that of individuals in household j. Even for such a simple generalization, specification of an appropriate correlation structure is difficult because of variable number of persons in each household. Even if one solves this problem, the household employment rate, in that case, may be sensitive to the functional specification of the joint probability distribution.

[4] In fact, in the concluding section of the Dissertation I argued for possible lack of independence across individuals and listed it as a potential future research problem.

The Implications of Non-stationarity on the Robustness of Regression Estimates

Referee 1 has wanted to know the exact consequence of the relaxation of the assumption of independence across individuals and also the implications of non-stationarity for the robustness of regressions estimates. Given the nature of the problem, this will be an extremely difficult task.

In this context, Referee 2 has observed that the finding of similar type of employment risk across regions might have occurred because of the short span of the reference period. This is, of course, a logical possibility. The average employment of individuals for seven days may not reflect the true steady state employment and can only be used as a very crude assumption. Hence, the regressions done in Chapter 6 may not yield the true coefficients. However, determination of the extent of robustness of the regression coefficients in this context will be too difficult and I can only express my helplessness.

A Logit Model for Village Employment Rate

The dependent variable village employment rate is bounded between 0 and 1. Since OLS does not take into account the boundedness of the dependent variable, Referee 1 has wanted to know why a probit or a logit model had not been estimated.

Following the advice of Referee 1, a logit model was estimated for village employment rate. The estimates of logit model have been presented in Table D.1.4. The sign of the coefficients are similar to that of the OLS model (except for Subround 4). The standard errors were calculated from the subsamples. Except the constant term and LABPC, all the other t-values turn out to be low.

- **The Specific Queries**

Page 10 : According to Appendix B, code 82 is excluded for classification according to usual status

Page 4 of the "Report on the Third Quinquennial Survey of Employment and Unemployment, Maharashtra" (National Sample Survey Organization, Government of India, 1988, No. 341/4) states that the activity category codes 61, 71, 82 and 99 were used only for current status approaches. Thus no individual is assigned the code 82 as usual status. An individual will be classified as unemployed for activity code 81 only. I thank Referee 1 for pointing out this error.

In spite of admission of my error, I feel some further clarifications are needed. NSSO assigns the codes separately for usual status, weekly status and daily status. Thus even if one falsely believes that code 82 pertains to usual status unemployment and counts it within his measure for usual status unemployment, he will not commit any mistake in calculations because he will not be able to trace an individual whose usual status is 82 !

In the Dissertation, I worked with measures based on daily status only. Since these codes are assigned separately, the reported calculations are not affected in any possible way.

 # Page 11 : Since codes 61-71 relate to an individual not working, it does not seem appropriate to include those codes in the definition of working for the entire day.

As in the previous case, individuals with the codes 61 and 71 were included as working because the NSSO definition includes them as working. Again, I quote from the same source : "Persons assigned any one of the activities listed under the category codes 01 to 71 were treated as 'working' (or employed)" (Page 5).

In the Dissertation, I followed the NSSO definition. This definition was followed for easy comparability. Such a comparison was presented in page 24 of my Dissertation. The slightly modified definition of employment adopted in the Dissertation yielded an unemployment rate of 6.75% vis-a-vis 6.62% of that of NSSO in rural Maharashtra.

 # The Specific Queries Regarding Page 65 and Pages 78-79

Referee 1 has pointed out two errors on my part. I am sorry for these errors that remained and offer the following corrections :

(a) On page 65, there is a notational mistake in Corollary 4.2.1. The notation adopted in it was not defined until page 68. For correct interpretation of the corollary, one should therefore use the notation adopted in Section 4.3 of the Dissertation.

(b) On pages 78-79, u must be defined as a column vector of K elements and x as a row vector so that $\mathbf{P} = ux'$ is a $K \times K$ matrix. Thus $u' = (1, 1, \ldots, 1)$.

REFERENCES

- **Coondoo,D., 1975 :** "Contributions to the Analysis of Consumer Behaviour", *Ph.D. Dissertation, Indian Statistical Institute.*

- **Deaton,A., 1994 :** "The Analysis of Household Surveys : Microeconomic Analysis for Development Policy", *Research Program in Development Studies, Princeton University, Draft.*

- **Fields,G.S. and Ok,E.A., 1996 :** "The Measurement of Income Mobility : An Introduction to the Literature" in 'Income Inequality Measurement : From Theory to Practice' by J. Silber [Ed.], *Kluwer Academic Publishers, Boston, Forthcoming.*

- **Gourieroux,C., Monfort,A. and Trognon,A., 1984 :** "Pseudo Maximum Likelihood Methods : Theory", *Econometrica*, **52**, 681–700.

- **Gourieroux,C., Monfort,A. and Trognon,A., 1984 :** "Pseudo Maximum Likelihood Methods : Applications to Poisson Models", *Econometrica*, **52**, 701–720.

- **Iyengar,N.S., 1960a :** "On a Method of Computing Engel Elasticities from Concentration Curves", *Econometrica*, **28**, 882–891.

- **Iyengar,N.S., 1960b :** "On a Problem of Estimating Increase in Consumer Demand", *Sankhya*, **22**, 379–390.

- **Iyengar,N.S., 1963 :** "Estimation of Quality Elasticities for Certain Commodities from National Sample Survey Data", *Sankhya*, **25**, 15–22.

- **Townsend,R.M., 1994 :** "Risk and Insurance in Village India", *Econometrica*, **62**, 539–591.

Note : Only the references which are not in the Dissertation have been given.

Table D.1.1 : Results of Kolmogorov-Smironov Tests of Homogeneity of Empirical Cumulative Distribution Functions

Null Hypothesis	Test Statistic	Critical Values 5% Level	1% Level
(1)	(2)	(3)	(4)
Region			
Coastal = Western*	0.0483	0.0318	0.0382
Coastal = Northern	0.1150	0.0348	0.0417
Coastal = Central	0.0485	0.0347	0.0416
Coastal = Ineast	0.1162	0.0343	0.0411
Coastal = Eastern	0.0679	0.0409	0.0490
Western = Northern	0.0667	0.0272	0.0326
Western = Central	0.0057	0.0272	0.0325
Western = Ineast	0.0679	0.0266	0.0318
Western = Eastern	0.0196	0.0347	0.0415
Northern = Central	0.0665	0.0305	0.0366
Northern = Ineast	0.0118	0.0300	0.0360
Northern = Eastern	0.0471	0.0374	0.0448
Central = Ineast	0.0677	0.0300	0.0359
Central = Eastern	0.0194	0.0373	0.0448
Ineast = Eastern	0.0483	0.0369	0.0442
Subround			
Subround1 = Subround2	0.0107	0.0245	0.0293
Subround1 = Subround3	0.0285	0.0258	0.0309
Subround1 = Subround4	0.0250	0.0246	0.0295
Subround2 = Subround3	0.0178	0.0258	0.0309
Subround2 = Subround4	0.0143	0.0245	0.0294
Subround3 = Subround4	0.0140	0.0259	0.0311
Occupation			
Selfagr = Selfnagr	0.0025	0.0338	0.0405
Selfagr = Agrlab	0.1539	0.0200	0.0240
Selfagr = Othlab	0.0969	0.0380	0.0455
Selfagr = Othocc	0.0163	0.0376	0.0451
Selfnagr = Agrlab	0.1519	0.0343	0.0411
Selfnagr = Othlab	0.0949	0.0471	0.0565
Selfnagr = Othocc	0.0183	0.0468	0.0561
Agrlab = Othlab	0.0570	0.0385	0.0461
Agrlab = Othocc	0.1702	0.0381	0.0456
Othlab = Othocc	0.1132	0.0499	0.0598

(Contd.)

* Here Coastal = Western means the null hypothesis that the cumulative distribution functions of these two regions are equal. The expressions in the subsequent rows are similar. The covariate and variable names are as in the Dissertation.

Table D.1.1 : Results of Kolmogorov-Smironov Tests of Homogeneity of Empirical Cumulative Distribution Functions (Concld.)

Null Hypothesis	Test Statistic	Critical Values 5% Level	Critical Values 1% Level
(1)	(2)	(3)	(4)
Social Group			
ST = SC	0.0134	0.0372	0.0446
ST = Othhind	0.0831	0.0262	0.0314
ST = Othmusm	0.0716	0.0476	0.0571
ST = Othgrp	0.0267	0.0412	0.0493
SC = Othhind	0.0697	0.0307	0.0368
SC = Othmusm	0.0582	0.0503	0.0602
SC = Othgrp	*0.0150*	*0.0442*	*0.0529*
Othhind = Othmusm	0.0115	0.0428	0.0512
Othhind = Othgrp	0.0681	0.0354	0.0424
Othmusm = Othgrp	0.0566	0.0533	0.0639
Age and Sex			
Boychild = Girlchild	0.0280	0.0307	0.0368
Adultmale = Adultfemale	0.0063	0.0227	0.0273
Oldmale = Oldfemale	0.0451	0.0756	0.0906
Boychild = Adultmale	*0.1554*	*0.0267*	*0.0320*
Boychild = Oldmale	0.0671	0.0579	0.0694
Adultmale = Oldmale	0.0883	0.0562	0.0674
Girlchild = Adultfemale	0.1304	0.0273	0.0327
Girlchild = Oldfemale	0.0060	0.0575	0.0689
Adultfemale = Oldfemale	0.1364	0.0554	0.0664
Education			
Illit = Primid	0.0537	0.0226	0.0271
Illit = Highed	*0.1416*	*0.0489*	0.0586
Primid = Highed	0.0879	0.0502	0.0601

Note : The covariate and variable names are as in the Dissertation.

Table D.1.2 : The Estimates of Logit Model of Employment Fluctuation : Rural Maharashtra, NSSO 38-th Round, Combined Sample

Coefficients	Subround 1	Subround 2	Subround 3	Subround 4
(1)	(2)	(3)	(4)	(5)
CONSTANT	-3.5730	-3.3010	-3.5920	-3.9330
	(0.043)*	(0.238)	(0.032)	(0.082)
WEST	-0.8022	-0.2960	-0.4049	-0.0894
	(0.259)	(0.003)	(0.213)	(0.096)
LABOUR	1.2780	0.9433	1.2950	1.3780
	(0.092)	(0.184)	(0.002)	(0.029)
SCST	0.0829	0.3798	0.2887	0.1871
	(0.190)	(0.041)	(0.222)	(0.179)
CHILD	-2.0580	-2.3280	-1.9140	-1.9770
	(0.263)	(0.097)	(0.233)	(0.079)
OLD	-0.8843	-0.6192	-1.2360	-0.9472
	(0.094)	(0.117)	(0.202)	(0.258)
FEMALE	0.1202	-0.3672	0.2990	0.3358
	(0.044)	(0.209)	(0.087)	(0.117)
HIGHED	-0.5383	-1.3470	-0.9828	-0.7722
	(0.139)	(0.158)	(0.114)	(0.306)
Log-likelihood	-4573.77	-5062.59	-4578.89	-5086.46
Sample Size	6162	6203	5032	6077

* The numbers in bracket are the estimated standard errors of the estimates.

Note : The covariate and variable names are as in the Dissertation.

Table D.1.3 : t-Ratio of Mobility of Various Covariate Groups Separately across Subrounds : *Rural Maharashtra, NSSO 38-th Round.*

Covariates	Categories	Subround 1	Subround 2	Subround 3	Subround 4
(1)	(2)	(3)	(4)	(5)	(6)
Region	Coastal	2.38	3.12	5.36	7.98
	Western	3.44	18.95	3.70	6.19
	Northern	4.52	13.56	9.44	16.14
	Central	34.64	3.93	2.42	555.00
	Ineast	8.43	36.52	18.60	10.84
	Eastern	19.08	1.64	2.43	8.61
Household Type	Selfagr	15.11	9.81	9.43	9.85
	Selfnagr	578.00	3.08	6.50	2.63
	Agrlab	9.96	32.00	12.72	94.26
	Othlab	2.86	18.19	5.70	8.77
	Othocc	5.31	8.70	5.35	9.56
Social Group	ST	8.68	157.57	11.20	20.02
	SC	4.49	27.25	4.11	7.26
	Othhind	13.23	8.77	8.21	11.87
	Othmusm	8.92	4.24	4.03	20.77
	Othgrp	3.71	1.72	9.05	11.44
Age and Sex	Boychild	4.31	28.17	2.12	2.05
	Girlchild	5.47	4.68	7.52	42.27
	Adultmale	104.93	9.78	6.90	14.37
	Adultfemale	9.63	9.48	13.15	37.00
	Oldmale	8.08	5.35	3.66	9.10
	Oldfemale	176.00	3.66	7.03	1.08
Education	Illit	17.76	19.79	14.01	35.59
	Primid	12.91	6.63	7.21	7.45
	Highed	9.48	4.87	4.02	5.66

Note : The covariate and variable names are as in the Dissertation.

Table D.1.4 : The Estimates of Logit Model of Village Employment Rate : *Rural Maharashtra, NSSO 38-th Round, Combined Sample*

Variables	Coefficient	Standard Error
(1)	(2)	(3)
CONSTANT	2.794	0.0650
WEST	0.416	0.2025
LABPC	-1.235	0.0165
SUBRND 2	-0.239	0.1010
SUBROUND 3	-0.060	0.0447
SUBROUND 4	-0.214	0.1156
LANDCULT	0.00058	0.0001
Log-likelihood = -19419.06, Sample Size = 540		

Note : The covariate and variable names are as in the Dissertation.