South African Unemployment

A Supply Side Analysis of the Labour Market

By

James R. Wilson

University of Cape Town

Dissertation submitted to the University of Cape Town in partial fulfilment of the Degree of Master of Social Science in Economics

September 1999

Abstract

This study of South African unemployment approaches the problem from the supply side of the labour market. A descriptive analysis of South African unemployment is made first of all, using data from the 1995 October Household Survey. This is followed by a discussion of various important elements of the South African unemployment debate. An introduction to some theoretical and econometric considerations around labour supply modelling is then given, leading into a statistical and econometric analysis of the African labour supply decision using OHS 1995 data. In particular the labour supply process is split into three stages - participation, ability to find employment, and hours - and each stage analysed in detail using appropriate techniques, including probit and Tobit models for dichotomous and censored dependent variables. In terms of participation, wages are found to have a considerably larger association with female participation than with that of males, and non-wage income shows a very small negative relationship for both groups. Education is found to have a much larger influence on the ability of female participants to find employment than males, and the ability for males to find employment is strongly associated with aspects of location. In particular, the Eastern Cape and Northern Province are identified as regions where the chances of participants finding employment are greatly reduced. Finally, overall labour supply elasticities are found to be positive and inelastic for both males and females of the African population group.

Acknowledgements

I would like to express special thanks to my dissertation supervisor, Murray Leibbrandt, for his time, guidance and valuable assistance throughout, and for always being positive and inspiring about research in this area. I would also like to thank Matthew Welsh at SALDRU for his time and help in familiarising me with the OHS database, and Brian Khan for easing some of the computer difficulties at UCT.

Thanks also to my friend David Crosoer for many valuable discussions and quiet drinks, and for printing the final copy for me in Cape Town. Finally, a very special thank-you to Miriam and to my parents for their huge support, love and encouragement from afar, and also to my grandparents here in Cape Town.

Contents

CHAPTER ONE: An Introduction to the Study **1**
1.1 Introduction 1
1.2 The Importance of Finding Solutions to South Africa's Unemployment 2
1.2.1 The Central Role of Unemployment to South Africa's Problems 3
1.3 The South African Debate and the Scope for Labour Supply Analysis 5
1.3.1 What Can an Analysis of Labour Supply Add? 6
1.4 The Structure of the Study 9

CHAPTER TWO: The South African Unemployment Debate **10**
2.1 Introduction 10
2.2 Debate over South African Unemployment Figures 10
2.3 Characterising South African Unemployment 13
2.3.1 Unemployment by Race 14
2.3.2 Unemployment by Location and Gender 15
2.3.3 Unemployment by Education 17
2.3.4 Unemployment by Age and Race 17
2.4 Elements of the Unemployment Debate 19
2.4.1 How do the Unemployed Survive? 19
2.4.2 The Urban-Rural Dichotomy 21
2.4.3 The Racial Divide 22
2.4.4 Issues of Age and Gender 23

CHAPTER THREE: Labour Supply Theory and Modelling **26**
3.1 Introduction 26
3.2 Neo-classical Labour Supply Theory 27
3.3 From Theory to an Empirical Modelling 29
3.3.1 Specifying the Model to Accounting for Different Individuals 29
3.3.2 Addressing the Non-Participation Problem 31
3.4 The Practicalities of Modelling Labour Supply 33
3.4.1 Sample Selection Bias 34
3.4.2 Dealing with a Censored Dependant Variable 35
3.4.3 Tobit Models of Labour Supply 36
3.4.4 Sample Selection Models of Labour Supply 37
3.5 Labour Market Rationing and the Need to Tailor Conventional Models 39
3.5.1 Criticisms of the Neo-classical Model 39
3.5.2 Unemployment as a Rationing Problem 40

CHAPTER FOUR: An Introductory Analysis of the Data **43**
4.1 Introduction 43
4.2 The Data Set 43
4.2.1 The OHS 1995 Survey 43

4.2.2 Sampling Issues 44
4.2.3 Identification of Labour Supply Variables 45
4.2.4 Customising the Data Set 46

4.3 A Descriptive Analysis of African Labour Supply 47
4.3.1 African Males 47
4.3.2 African Females 50
4.4 Awareness of the Potential for Measurement Errors 53

CHAPTER FIVE: An Econometric Model of African Labour Supply 55
5.1 Introduction 55
5.2 Wage Equations 56
5.3 The Participation Decision 60
5.4 The Ability to Find Employment 64
5.5 Estimating Hours Models of Labour Supply 67

SUMMARY AND CONCLUSIONS 73

BIBLIOGRAPHY 81

APPENDICES 86
Appendix One 86
Appendix Two 88
Appendix Three 89

List of Figures and Tables

FIGURES

Figure 3.1: Determination of Desired Hours 28
Figure 3.2: Negative Desired Hours 31
Figure 4.1: The Distribution of Weekly Hours (Males) 49
Figure 4.2: The Distributions of Log Weekly and Log Hourly Earnings (Males) 50
Figure 4.3: The Distribution of Weekly Hours (Females) 52
Figure 4.4: The distributions of Log Weekly and Log Hourly Earnings (Females) 52

TABLES

Table 2.1: South African Unemployment Rates (1995) 14
Table 2.2: Unemployment Rates by Race (1995) 15
Table 2.3: Unemployment Rates in Urban and Rural Areas (1995) 15
Table 2.4: Provincial Unemployment Rates (1995) 16
Table 2.5: Unemployment Rates by Educational Achievement (1995) 17
Table 2.6: Unemployment Rates by Age Group (1995) 18
Table 2.7: Unemployment Rates by Race and Age Group (1995) 18
Table 4.1: Mean and Standard Deviations of Key Variables (Males) 48
Table 4.2: Mean and Standard Deviations of Key Variables (Females) 51
Table 5.1: Wage Equation Estimates for African Males 57
Table 5.2: Wage Equation Estimates for African Females 58
Table 5.3: Summary Statistics of Predicted Log Wage (Males) 59
Table 5.4: Summary Statistics of Predicted Log Wage (Females) 59
Table 5.5: Summary Statistic Comparisons for Imputed Wage Variables 59
Table 5.6: Probit Estimates of the Participation Decision 62
Table 5.7: Probit Estimates of the Ability of Participants to Find Employment . 66
Table 5.8: Estimates of Labour Supply Equations (Males) 69
Table 5.9: Estimates of Labour Supply Equations (Females) 70

Chapter One

An Introduction to the Study

1.1 Introduction

One of the most striking features of the world economy at the end of the 1990s, and indeed as we look back across most of the last century, is the large amount of unemployment and underemployment that can be observed.[1] In developed and developing nations alike, the problem of unemployment has consistently plagued society, and has proven itself to be the most pervasive of economic ills. Here in South Africa unemployment is of particular concern as we reach the end of the twentieth century, with survey data consistently estimating broad unemployment in the region of thirty percent overall, and much higher among disadvantaged population groups.[2] Unsurprisingly there has been much debate in South Africa over the causes and the consequent solutions to this mass unemployment, and in this study I aim to contribute to the debate through an analysis of the supply side of the African labour market.

In this introductory chapter I emphasise the importance of finding better solutions to unemployment, focusing specifically on the significance of unemployment in the economy of South Africa. This discussion highlights some of the key issues in the ongoing unemployment debate in South Africa, and in this context I introduce the analysis of labour supply as a potentially important contributor to the debate. Finally, I conclude the chapter by outlining the structure in which the study will proceed.

[1]As an indication of the extent of the global problem, a 1996 report by the International Labour Organisation (ILO) estimated that world unemployment had reached one billion, with almost a third of the global workforce unemployed or underemployed [The Guardian (Nov. 26 1996)].

[2]Klasen & Woolard (1999:11, table 1) report broad overall unemployment rates of 32.6% and 29.3% for 1994 and 1995 respectively, calculated using OHS data and by the official Central Statistical Services (CSS) definition of unemployment.

1.2 The Importance of Finding Solutions to South Africa's Unemployment

Unemployment is arguably the most important economic problem of all, and yet it remains an area of considerable mystery, where the shortage of adequate theory and policy is well acknowledged. Unemployment is a crucial issue in all societies, not only because it represents an inefficient use of resources, but because it undermines the ability of many people to earn a sufficient living, and to feel a pride and contentment with their place in society.[3] From an efficiency perspective, unemployment results in a lower national output than is potentially possible, and generates poor returns on public (and private) investments that have been made on the education and training of those that are unemployed [McLaughlin (1992:4)]. From a personal and social perspective, it jeopardises people's ability to support themselves and their families, and creates divisions in society between those with and those without jobs.

The phenomenon of unemployment is therefore inextricably linked to important issues such as poverty, inequality, crime, growth, and general economic and social well-being. Joblessness effects people's lives in the most fundamental way, and it is for this reason that the study of unemployment, and the development of policies to generate meaningful and lasting employment, is so important.

For low-income economies, future development is dependent more than ever on their initial ability to create greater employment opportunities.[4] Recent years have seen a marked restriction on the ability of developing countries to access credit, and with increased mobility of capital, international markets treat severely countries whose internal and external debt are perceived to be unsustainable. Previous channels of funding for infrastructure development, education, and basic living standards provision are therefore restricted in many cases, and the importance of generating jobs to improve government

[3]Some people may indeed be very happily unemployed, but in general it is fair to assume that most would prefer suitable employment. The economic psychology literature has produced a number of studies, for example, which aim to examine the determinants of some notion of individual utility or happiness. Such studies in Western economies have generally found that unemployment has an important negative influence on measures of well-being. See, for example, Clark & Oswald (1994).

[4]In many low-income countries the phenomenon of open unemployment is in fact relatively rare, but the problem lies instead with severe under-utilisation of labour, which prevents large groups from earning enough to lift themselves from poverty [World Bank (1995:27)].

finances is heightened. Without creating initial employment opportunities, it will be difficult for most developing countries to generate sufficient funds to improve infrastructure, education, and basic living standards for the general population. These are all factors that should in turn stimulate further investment, growth and jobs, but it is vital that the most effective policies are found to create jobs in the first instance. In developing countries in particular, therefore, employment creation should be seen as a priority.

1.2.1 The Central Role of Unemployment to South Africa's Problems

Turning to South Africa, we have an example of a country for which job creation really does hold the key to future success. Unemployment is high in South Africa by any standards, and while there is ongoing discussion concerning the accuracy of the official figures, there is little dispute that unemployment is the primary problem facing the country. A reflection of this concern can be seen in the campaigns for the recent general elections, which have seen all parties strongly acknowledge the importance of job creation to the future of South Africa.

If we consider the main economic and social issues in South Africa at present, then it becomes clear that unemployment is fundamental to all of them. Poverty, first of all, is a huge problem in South Africa, and effects a large proportion of the population, and an especially large proportion of those living in rural areas. In 1993, for example, almost 24 percent of the total population were below the international poverty line of $1 per day at 1985 prices. With more generous poverty lines, this percentage rises to between 45 and 57 percent [Leibbrandt & Woolard (1999:40, table 1)].

While the incidence of poverty does not necessarily correspond to the incidence of unemployment, there is likely to be a strong link between the two. The ILO Country Review of South Africa argues that the limited evidence suggests a link between unemployment and poverty "but that the degree of overlap is far from complete" [ILO (1996:110)]. However, in a detailed analysis of OHS and IES data from 1995, Klasen & Woolard (1998:15) show that "unemployment is closely related to poverty, with unemployment rates at 59% among the poorest quintile, compared to only 5% among the

richest".[5] On this later evidence the link would appear to be a strong one, and this underlines the importance of providing suitable employment opportunities if large numbers of people are to be lifted from poverty.

Connected to poverty is the issue of economic inequality, and this is a topic of particular sensitivity given South Africa's apartheid past. South Africa possesses some of the highest income inequality in the world, and with a Gini coefficient of around 0.6 it ranks alongside the most unequal of South American countries. Income inequality reflects a number of factors, but a recent decomposition study of the Gini coefficient has shown that wage income is the most important component of inequality among the African population [Leibbrandt, Woolard & Woolard (1996)]. Given the high rate of unemployment (especially among this population group), it is reasonable to assume that unemployment therefore plays a vital role in generating the unequal distribution of income in South Africa.

The related issues of poverty and inequality are important in their own right, but given the political history of South Africa they become doubly so. It is no longer acceptable to see such income gaps between races, and no longer acceptable to maintain large numbers of the African population in poverty. Not surprisingly, these are extremely sensitive issues, and addressing them successfully will be crucial for the social and political stability of a country that has only just had its second set of democratic elections. The political sensitivity surrounding the delivery of better living standards to the majority of the population, and the evening of the wealth distribution, therefore makes the challenge of addressing unemployment all the more important. Indeed, "there is a consensus that the failure of the economy to generate large-scale and sustained increases in employment undermines the possibility of social progress" [MERG (1993:152)]. Political and social stability are vital to create a conducive environment for both domestic and foreign investment, and in a sense this means that we have to create jobs in order to generate stable conditions for future growth and job creation.

[5]Furthermore, households in the bottom two consumption quintiles account for 64.6 percent of individuals who are unemployed at a broad measure, and 58.5 percent of those unemployed at a narrow measure [Klasen & Woolard (1998:16, table 7)].

A further disturbing consequence of unemployment, poverty and inequality is the massive crime rate that South Africa has been experiencing in recent years. Crime of many forms has escalated, and not only does this create a general social malaise, but it also harms the economic environment in a number of ways. Crime is an inefficient and socially damaging method of redistributing income, and its influence on domestic and foreign investment, and on important industries such as tourism, can only be negative. It has also arguably been one of the key reasons for the so-called 'brain drain', whereby many of South Africa's most skilled people have been leaving the country. Once again, logic would suggest that a natural way to reduce much of the crime in South Africa is to create decent employment opportunities, particularly for the youth of the country.[6]

1.3 The South African Debate and the Scope for Labour Supply Analysis

The links between unemployment and a variety of pressing social and economic problems in South Africa are indisputable, and the importance of finding solutions to unemployment is generally recognised in this regard. However, as emphasised in the ILO country report on South Africa (1996:128), "it cannot be presumed that just because unemployment is probably a manifestation of poverty and inequality that all policies designed to reduce unemployment would reduce poverty and inequality". While it is clear that unemployment must be tackled as a matter of paramount importance, great care must be taken in assuring that policies are selected which allow the reduction of unemployment to positively impact on poverty and inequality.

Indeed, it is this issue which forms the basis for much of the debate surrounding the unemployment problem in South Africa. In particular, debate has been focussed on the extent to which unemployment is structural or neo-classical in nature, and hence issues of labour market flexibility. This distinction is particularly important in selecting correct

[6]Admittedly this will not solve South Africa's crime problems alone, and may have limited influence in the short term. Indeed, there is a much larger debate to be had surrounding the whole crime issue. However, it is fair to say that other policies to reduce crime can only be effective in the longer term if the root causes of crime are eliminated. Unemployment and lack of opportunities (and their influence on poverty and inequality) are undoubtedly one of these root causes, and in the long term they are therefore central to the crime problem.

policies, and is closely linked to arguments surrounding the inequality and poverty consequences of certain measures. For example, a policy which encouraged falling wages to enhance employment may have the simultaneous effect of bringing those already only earning a subsistence wage into poverty, and at the same time increasing inequality by permitting a larger gap between the top and bottom of the earnings distribution.

Consequently there has been much discussion over the relative merits of improving labour market flexibility, of implementing policies to enhance skills and training (and hence tackle structural mis-match), and of the possibility of utilising more innovative and guided industrial policy measures to generate employment. Much of this policy debate has centred around the Government's 1996 macroeconomic policy document, GEAR, which outlines the intended policy approach towards the labour market [Ministry of Finance (1994:16-20)]. In particular, it has been criticised for relying too much on wage moderation and increased foreign investment for employment creation, and being quick to "downplay the leading role that Government could play in this regard" [NIEP (1996:18); see also Michie (1997:159-162)].

More specific issues encompassed in the general unemployment debate include the coping mechanisms of the unemployed, the observed difference in unemployment rates between urban and rural markets, the concentration of unemployment among the young (and particularly young Africans), and the linkages between education, training, and unemployment. Some of these issues and the debate surrounding them are discussed in more detail in the following chapter, in the context of a detailed statistical breakdown of the incidence of unemployment.

1.3.1 What Can an Analysis of Labour Supply Add?

While current economic theory provides a basic framework for thinking about the likely causes of unemployment and the available policy options, it is extremely limited in it's ability to prescribe concrete and generally applicable answers to the problem. Policies that might be effective under one set of circumstances may prove largely unsuitable in another. For example, the flexible labour market policies that have shown success in the

US are arguably unsuitable in their extreme form to the economy of South Africa at the present time. There is a strong recognition that previously enforced disparities and inequalities among population groups must be addressed, and one consequence of the free market is that it tends to reward those that already have, and thus widen the already present inequalities.

It is imperative, therefore, that policy responses to South African unemployment should be designed specifically with the concerns and needs of the South African economy in mind. In practice, this requires a fundamental study of the labour market, and the outcomes that it produces. We need to know what sort of people have jobs, what sort of people are looking for jobs, and what sort of people are not looking for jobs. We can then begin to make a clearer distinction between voluntary and involuntary unemployment, and identify rigidities and problems that are specific to the South African labour market.

The unemployment debate itself is naturally focused very much on the labour market, and in particular on aspects of flexibility and structural mis-match. There has indeed been much work done on different aspects of the labour market, with the general aim of identifying it's weaknesses, and making the policy debate around unemployment more informed.[7] However, there is a lot still to be gained with respect to these issues by taking a labour supply perspective, and it is from this angle that I aim to contribute to the South African unemployment debate.

Many people argue, and rightly so, that labour demand is vital for higher employment, and that therefore it is important to create conditions in the economy that will encourage firms to employ more people. However, there is little point in analysing labour demand without a corresponding understanding of labour supply. It is the meeting of the two forces that determines employment, and the nature of the labour supply decision is a crucial part of this.

Specifically, the study of labour supply enables us to answer important questions regarding the decision to participate in the labour market, the ability of participants to find

work, and the decision of how many hours of labour to supply. The second stage of this process is perhaps the most important, as this is where the forces of demand and supply meet in determining whether somebody who desires a job can find one. The analysis of this three-stage process allows us to investigate factors that are important in the decisions that potential workers make, and how these interact with demand in the market place. This can potentially give us a clearer picture of how the labour market works, and therefore how demand policies can be most effectively directed, and complemented by appropriate supply-side policies.

How responsive are labour participation and supply to wage rates and other factors? And how does this vary by characteristics such as education, age, race, gender or region? There are likely to be large differences, for example, between female and male labour supply, and between supply in rural and urban labour markets. Furthermore, what determines the distinction between participation and employment? These are questions that need to be considered, as such issues will influence the effectiveness of certain policies to reduce unemployment, and guide us to specific areas where attention is needed. By studying the process of labour supply in these stages I hope to distinguish between factors of importance that influence participation, and then those that in turn influence the ability of participants to find employment. Also of interest are hours elasticities, as these give some indication of the responsiveness of labour supply overall to various factors, and are potentially crucial in the debate over the respective merits of different policy responses.

In particular, this study will focus on the labour supply process for the African population group, as it is here that unemployment rates are especally large. Given that Africans constitute around 70% of the potential labour force, it is indeed clear that solutions to the overall problem lie with the labour market problems of this population group.

[7]For example, Hofmeyr (1998), Klasen & Woolard (1998, 1999), ILO (1996), Schultz & Mwabu (1998),

1.4 The Structure of the Study

The remainder of the study is structured as follows. In the next chapter I examine in more detail some important facets of the unemployment debate in South Africa. The chapter starts with a brief discussion of the measurement of unemployment in South Africa, and then proceeds with the presentation of some statistics to paint a fairly detailed picture of the unemployment situation as it existed in 1995. Various facets of the unemployment debate are then discussed in the context of these figures.

In the third chapter I move to the specific topic of labour supply, and introduce the theory on which the econometric models of labour supply are based. This chapter traces the formulation of labour supply models from neo-classical theory, discussing the various difficulties that arise and the econometric solutions offered. It arrives at a series of conventionally used approaches for modelling the labour supply decision, and then briefly critiques these approaches and suggests their careful use in the study of the labour supply decision in a heavily rationed economy such as South Africa.

The next two chapters constitute the analytical part of the study, with chapter four providing an introductory analysis of the data, and chapter five the econometrics. Chapter four starts with an introduction to the October Household Survey (OHS) dataset, and then presents summary statistics and discussion for key variables that are thought to determine wages, the participation decision, the ability to find a job, and the hours of work that are ultimately supplied. This is confined to the population group of interest, the African population, and split by gender.

Chapter five then utilises some of the econometric techniques discussed in chapter three to model the three elements of the labour supply decision in a step-by-step approach. The decision to participate, the ability of participants to find a job, and the decision of how many hours to supply are therefore analysed separately to build a complete picture. A summary of the findings and drawing together of different ideas within the various chapters is then made to conclude the study.

Winter (1998), Wittenburg (1999).

Chapter Two

The South African Unemployment Debate

2.1 Introduction

Given the importance of unemployment to so many of South Africa's problems and the great challenge that tackling unemployment itself poses, it is unsurprising that there has been significant debate around the issue in South Africa. This debate has taken place among economists, politicians, and the general population at large. Indeed, the recent election campaigns showed a significant focus on unemployment, and saw each party emphasising the importance of job creation. While the various stances on how this should be approached have displayed common themes, there are also noticeable divergences in opinion. The debate is certainly varied, with views differing not only on the causes of the problem and how to tackle it, but also on the magnitude of the unemployment figures themselves.

In this chapter I firstly address some of the issues surrounding the measurement of unemployment in South Africa and the reliability of various figures. I then set about describing in detail the unemployment situation, and examining how unemployment is distributed among the population. In the final section I move on to review the current discussion surrounding certain facets of South African unemployment and their causes.

2.2 Debate over South African Unemployment Figures

South Africa has one of the highest reported unemployment rates in the world, and with such extreme figures there has been considerable concern over their calculation and reliability. There are various sources of labour market statistics in South Africa, the primary ones being the Central Statistical Service (now re-named Statistics South Africa)

and the Department of Labour.[1] From these sources a variety of measures of unemployment emerge, giving a significant range of different estimates of the unemployment rate. It is vital therefore to express care when considering measures of unemployment, and to question what conceptualisation of unemployment different figures reflect.

Specifically, there are various different types of unemployment, each with slightly different causes and motivations. Such complexity means that different measures of unemployment may therefore capture different elements of the problem. We must be particularly careful, for example, to distinguish between voluntary and involuntary unemployment if figures are to be useful for policy analysis. Indeed, as noted in the ILO country report on South Africa (1996), one of the reasons why unemployment is difficult to define is because:

> "it combines a *condition* (being without employment), a *need* (for work or for income), an *attitude* (desire for paid work), a *capacity* (ability to accept an opportunity, or at least availability to do so), and an *activity* (searching for work)" [ILO (1996:103)].

Different interpretations of the above will result in broader or narrower reflections of unemployment, where broader measures allow increasing scope for 'discouraged workers' to be included. Different unemployment measures may also be more or less effective at highlighting the frictional, cyclical and structural elements of unemployment.

The narrowest measure of unemployment in South Africa is the percentage of the labour force who have officially registered as unemployed, for purposes of claiming benefits from the Unemployment Insurance Fund (UIF) or making use of the job-finding services offered by the government. This figure is extremely restrictive however, because the UIF has very low coverage and the extent of job-finding services are limited [Klasen & Woolard (1999:7)]. Consequently, the registered figure will seriously underestimate the unemployment problem. Of the other more useful measures, the most commonly used are

[1] For a full discussion of the most commonly used sources, and some criticisms aimed at the reliability of

probably the figures calculated from the CSS October Household Survey (OHS). As I will be using OHS data in this study, I will focus on these figures for the remainder of the discussion.

The CSS calculates measures of unemployment from the OHS using two definitions; a strict definition and a broader definition.[2] Conceptually, the difference between the two is that the broader definition includes those 'discouraged' workers who desire employment but have given up looking for work. Klasen & Woolard (1999:11, table 1) report figures of 20.3% and 16.7% for 1994 and 1995 respectively, calculated using the strict CSS definition for the total population aged 16-64. The broader CSS definition gives much higher rates of 32.6% and 29.3% in respective years.

The ILO was critical of the CSS definitions in its 1996 report, however, maintaining that "the strict definition is not as strict as in many countries and the broad is very broad indeed" [ILO (1996:108)].[3] There were also criticisms levelled at the survey itself. In particular there is concern that it missed a significant amount of workers living in mining hostels, defence force camps, hospitals and hotels, through a systematic under-sampling of these places of residence [ILO (1996:70)]. It is certainly possible therefore that these figures overstate the extent of unemployment in South Africa, both in terms of problems with the underlying data collection, and through being inconsistent with many international unemployment definitions.

Following the ILO report and other criticisms, Klasen and Woolard (1999) make a detailed analysis of the levels, trends and consistency of various employment and unemployment figures in South Africa. They find that despite the above criticisms, the two OHS surveys and the SALDRU household survey of 1993 offer a consistent picture of employment levels and trends. They conclude therefore that "the generally bleak picture of very high unemployment rates presented in the household surveys is broadly accurate" [Klasen &

---

certain sources see Klasen & Woolard (1999:2-4).

[2] To be strictly unemployed by the CSS definition, an individual has to be willing to accept a suitable job if offered, and have actively sought employment in the previous four weeks (corresponding to questions 3.28 and 3.30 in the OHS 1995 questionnaire). The broad definition drops the requirement that the individual must have been seeking work. See Klasen & Woolard (1999), ILO (1996).

[3] See ILO (1996) for a more detailed discussion of specific issues.

Woolard (1999:3)]. This is a disturbing conclusion in that it confirms the magnitude of the unemployment problem in South Africa. However, it does give us reason to be relatively confident in the OHS figures and, while caution is still needed, it opens the way for using OHS figures to analyse the problem in detail.

2.3 Characterising South African Unemployment

Unemployment does not occur randomly in society, and we would expect groups of individuals with certain characteristics to be effected in different ways and to varying degrees. Gender, race, age, location and education are all factors that are likely to be associated with vastly different rates of employment and unemployment, and inequities among groups may have important social and political consequences. It is important to know something about the incidence of unemployment therefore, and this is particularly so in South Africa, where certain groups have been greatly discriminated against in the past. By studying rates of unemployment among different groups we can also hope to learn something about the nature and functioning of the labour market and, importantly, this may help us to target policies for reducing unemployment in the most effective and efficient way. In this section I will briefly summarise the unemployment situation in South Africa, with the aim of identifying where in particular the labour market is struggling to provide employment.

Using data from the OHS 1995 survey I have classified individuals as broadly or strictly unemployed, based on a careful study of the OHS questionnaire. The basic principles of my classification are the same as those of the CSS unemployment definitions, and hence the results should be broadly similar. Generally, an individual is unemployed by the strict definition if he/she both desires work, and has taken active steps in seeking work in the last four weeks.[4] The broad unemployment classification drops the last condition, and thus includes those discouraged workers who have given up looking for work.[5] While

[4] As noted in the ILO report (1996:108) this is broader than a corresponding definition in many other countries, which would require active steps to have been taken in the reference week. This cannot be adjusted *ex-post* however, as it is inherent in the design of the questionnaire.

[5] The ILO report (1996:125) makes a valid point warning against a simple interpretation of such workers as discouraged workers, because "one does not know whether they have ever looked for work let alone

conceptually the same, my classifications differ slightly from those of the CSS in the way that they treat certain responses, and consequently the figures will be different to some extent.[6]

Table 2.1 South African Unemployment Rates (1995)

	Broad Unemployment	*Strict Unemployment*
Male	22.6	12.7
Female	37.6	20.7
Total	29.2	15.9

Source: OHS 1995

Table 2.1 gives broad and strict unemployment rates for the male, female and total populations, where the unemployment rate is equal to the number of unemployed individuals divided by the total number of labour market participants.[7] The figures are certainly high, with almost 16% of those who are actively in the labour force being unable to find work. Furthermore, the magnitude of the discouraged worker effect is striking, and the unemployment rate almost doubles to 29.2% when those who are not actively seeking work are included. The figures in table 2.1 also highlight a significant difference between male and female unemployment rates, with a much larger proportion of female participants by both measures being unable to find employment.

2.3.1 Unemployment by Race

Unemployment rates in South Africa vary dramatically across the four commonly used racial classifications, and the figures in table 2.2 are indeed quite staggering. Among the African population over 36% of broad participants are without work, and the strict unemployment rate is almost 20%. This is in contrast to 3.3% strict unemployment among whites, a remarkably low rate that compares favourably to almost any reported unemployment in the world! The coloured and Asian groups lie in-between, and it is noticeable among these groups (and also the white group) that narrow unemployment is

whether they have been discouraged by the experience". While it may not be strictly correct, I will continue to use this term to refer to those who desire a job but are not actively searching.

[6] In particular, they have incorporated some of the adjustments made to the CSS classifications by Klasen & Woolard (1999:10-11). See *Appendix 1* for a summary of precisely how the classifications were made.

[7] Labour market participants are those who are either employed or unemployed by the specific classification in question. Thus the strict unemployment rate only considers the strictly unemployed as participants, and not those who are just unemployed by the broad definition.

significantly larger relative to broad unemployment than for the African group or indeed the overall figures. This would maybe indicate that the discouraged worker phenomenon is especially large among the African population, something that is perhaps unsurprising given the high unemployment rate among this group, and the fact that low chances of finding work tend to breed discouragement.

Table 2.2 Unemployment Rates by Race (1995)

	Broad Unemployment	*Strict Unemployment*
African	36.7	19.9
Coloured	22.3	15.7
Asian	13.1	9.8
White	4.5	3.3

Source: OHS 1995

It is certainly clear from these figures that the unemployment problem in South Africa lies predominantly with the African population group, who constitute around 70% of the potential workforce. The key to addressing unemployment in South Africa must therefore lie with the labour market problems of this group.

2.3.2 Unemployment by Location and Gender

Examining the figures by location also tells us something important about the incidence of unemployment, and from table 2.3 the problem appears to be more severe in rural than in urban areas.

Table 2.3 Unemployment Rates in Urban and Rural Areas (1995)

	Broad Unemployment		*Strict Unemployment*	
	Rural	*Urban*	*Rural*	*Urban*
Male	28.0	19.7	13.0	12.5
Female	49.2	30.3	25.2	18.3
Total	37.8	24.2	18.1	14.9

Source: OHS 1995

The figures indicate a noticeable reduction in the ability of rural inhabitants to find work when compared with their urban counterparts. Just over 18% of those actively seeking work are unemployed in rural areas, as opposed to just under 15% in urban areas. More striking however are the levels of broad unemployment, where the gap widens considerably to indicate that there is a significantly larger discouraged worker effect in

rural areas. Table 2.3 also breaks the figures down by gender, and indicates that rural unemployment is a particularly large problem amongst females, although in general the gender dynamics broadly follow the pattern among the overall population.

Table 2.4 Provincial Unemployment Rates (1995)

	road Unemployment	*Strict Unemployment*
Western Cape	18.4	12.7
Eastern Cape	42.5	23.6
Northern Cape	27.3	17.9
Free State	25.9	11.4
Kwa-Zulu Natal	33.4	19.9
North West	33.2	16.6
Gauteng	19.6	12.2
Mpumalanga	34.2	15.5
Northern Province	41.4	18.5
All South Africa	29.2	15.9

Source: OHS 1995

South Africa is divided into nine provinces, and in table 2.4 I have presented the unemployment rates in each of these. Broad unemployment is lowest in the Western Cape and Gauteng, an unsurprising finding given that these two provinces are dominated by large metropolitan areas and are the wealthiest in South Africa.[8] Unemployment is highest in the Eastern Cape and the Northern Province, which are both predominantly rural areas and are similar in containing large parts of the former TVBC states.[9] Indeed, with broad unemployment of over 40%, these two provinces can be identified as areas where particular attention is needed. The strict unemployment rates follow similar patterns, although the Free State exhibits the lowest rate with just 11.4%. This suggests a relatively high proportion of discouraged workers to active job seekers in this province,[10] whereas in the Western Province and Gauteng the discouraged worker effect is comparatively low.

[8] Mean Monthly Earnings (over the total working age population) in Gauteng and the Western Province are R1864 and R1247 respectively. This is compared to an average of R944 across the whole country, and R509 in the Eastern Cape. These figures are calculated from the OHS 1995 data.

[9] The TBVC states refer to the qu asi-independent African homelands of the Transkei, Bophuthatswana, Venda and the Ciskei, which were created under apartheid policies.

[10] This pattern is also evident in other predominantly rural provinces such as North West and Mpumalanga.

2.3.3 *Unemployment by Education*

Different levels of educational achievement are also likely to be associated with different levels of unemployment, and we can see from table 2.5 that this is certainly the case in South Africa.

Table 2.5 Unemployment Rates by Educational Achievement (1995)

	Broad Unemployment	*Strict Unemployment*
No Education	32.1	13.2
Primary Education	36.0	18.7
Some Secondary Education	33.7	19.9
Education to Matric (18)	24.8	15.3
Tertiary Education	5.5	3.4

Source: OHS 1995

Generally we would expect employers to select the most educated and skilled individuals into employment, leaving those without education extremely vulnerable to unemployment. This pattern is evident in South Africa to a certain extent, but interestingly those without any education show lower unemployment rates than those with some primary and secondary education. Furthermore, the strict unemployment rate among individuals with matric is also higher than among those with no education, although the rate drops markedly as we move up to the tertiary education category. This pattern is surprising, and in the first instance prompts us to ask questions about how employers select their employees, and in particular about the quality of education that there is to choose from, and the type of labour demand in the economy (largely unskilled?). The broad unemployment figures follow a more expected pattern than the strict figures however, suggesting that the less educated are relatively more likely to stop looking for employment than the more educated. In turn, this goes some way to explaining the observed pattern in the strict figures.

2.3.4 *Unemployment by Age and Race*

Another important aspect of the incidence of unemployment is its distribution across age cohorts in the economy. We can see a distinct pattern from the figures in table 2.6, where

both measures of unemployment increase rapidly as we move into younger age groups. Strict unemployment is just 5.3% for those aged between 55 and 64, whereas for 16 to 24 year-olds it is a massive 32.9%. Broad unemployment follows a similar pattern, peaking at 50% for those in the youngest age group. Unemployment rates appear relatively stable between the ages of 35 and 54, however, and most noticeable is the massive increase in unemployment as we move to the 25-34 age group, and then again to the 16-24 group.

Table 2.6 Unemployment Rates by Age Group (1995)

	Broad Unemployment	*Strict Unemployment*
Aged 16-24	50.1	32.9
Aged 25-34	32.9	18.6
Aged 35-44	21.3	10.6
Aged 45-54	17.6	8.0
Aged 55-64	13.1	5.3

Source: OHS 1995

Table 2.7 Unemployment Rates by Race and Age Group (1995)

	Broad Unemployment				*Strict Unemployment*			
	African	*Coloured*	*Asian*	*White*	*African*	*Coloured*	*Asian*	*White*
Aged 16-24	64.7	38.3	24.4	10.6	44.5	29.6	19.7	8.8
Aged 25-34	41.3	20.4	10.7	3.1	23.9	14.1	7.0	2.7
Aged 35-44	26.8	14.6	9.3	3.0	13.3	9.5	7.1	1.6
Aged 45-54	22.8	13.0	5.7	3.9	10.0	8.9	4.4	2.4
Aged 55-64	16.7	7.8	16.0	3.5	6.5	3.3	12.4	1.8
Total	36.7	22.3	13.1	4.5	19.9	15.7	9.8	3.3

Source: OHS 1995

Table 2.7 illustrates the age profile now split by race, and we can see that patterns of unemployment across age groups do show certain differences among the races. In particular, white unemployment is stable at a very low level among all people above the age of 25, and then jumps quite dramatically for the youngest age group, indicating that young whites are also being effected by the youth unemployment problem.[11] A similar story (but at higher rates) is true of Asian unemployment, although there is an interesting leap in unemployment above the age of 55. A possible explanation for this may be a lower propensity to retire among this population group when work is not forthcoming. The trend among Africans and coloureds very much mirrors the overall trend by age, but at lower levels for coloureds and higher levels for Africans. Once again we see the

significance of the African unemployment problem, standing out above all other groups. Youth unemployment in a broad sense is almost 65% for Africans, and of those who are actively seeking work nearly 45% cannot find a suitable job.

2.4 Elements of the Unemployment Debate

The above characterisation of unemployment in South Africa gives us certain clues as to where the labour market is failing to provide jobs. We can see that unemployment among the African population group is of particular concern. Indeed, given that Africans make up 70% of the potential labour force, it is clear that unemployment in South Africa is predominantly an African phenomenon. Furthermore, the statistics suggest that unemployment is also especially prevalent amongst females, young people, and those in rural areas. These facets of the incidence of unemployment, as well as the overall magnitude of the problem, have stimulated discussion in a number of specific areas, and in this section I will examine some of the issues around which the South African unemployment debate is focussed.

2.4.1 How do the Unemployed Survive?

One of the first questions that springs to mind when confronted with such high unemployment figures, especially concentrated among one population group, is how such a large number of unemployed manage to support themselves. Indeed the OHS 1995 data indicates that only 41% of the working age population are actually engaged in employment, and this falls to only 35% for the working age African population. That this remarkable situation can arise when unemployment benefits are also scarce certainly requires some explanation.[12]

[11] This is also indicative of whites in older age groups being able to hold on to jobs that were previously provided (almost guaranteed) under apartheid.

[12] Only around 3% of South Africa's unemployed are receiving support at any one time [Klasen & Woolard (1998:3)].

There are suggestions that to some extent this must indicate an underestimate of informal activity, and that many of the unemployed must have some informal earnings. However, Klasen & Woolard (1998) find evidence in the survey data which lends strength to the idea that a large proportion of the unemployed are in fact supported by other members of the household. They address the question by asking what types of households the unemployed live in, and find that around 60% live in households where another member is employed. Of the remaining 40% only about half live in households with no connection at all to the labour market, the rest living in households that receive remittances from other households [Klasen & Woolard (1998:12)].[13] It is argued that "this outcome must be related to household formation strategies that ensure that unemployed persons are attached to a household where other members are employed, or receive other forms of income" [*Ibid.*:16].

The evidence suggests therefore that a major coping strategy of the unemployed revolves around the adapting of household structures, and this is an important point to consider when evaluating appropriate policy responses to unemployment.[14] It may be the case, for example, that the unemployed actually impart an upward pressure on wages, through their effect on raising the wage necessary for the household wage earner(s) to support the subsistence needs of the household [ILO (1996:113)]. In this case, a policy which resulted in falling wages could seriously harm many of the unemployed who still fail to find a job (and those that support them), and draw such households even deeper into poverty.

Furthermore, there have also been suggestions that coping strategies such as these can alter search behaviour, and in particular drastically lower the chances of employment if the unemployed have to withdraw to rural areas to obtain economic support [Klasen & Woolard (1998)]. This argument is closely linked to another interesting facet of South African unemployment, the urban-rural differences in unemployment rates.

[13] Furthermore, they find that only 12.6% of households contain no employed person and receive no remittances. This is made up partly of predominantly white households who rely on private pensions or incomes, and partly of African households who are struggling to survive on a combination of social grants, and some agricultural or minor self-employment income [Klasen & Woolard (1998:12-15)].

2.4.2 The Urban-Rural Dichotomy

Urban unemployment in South Africa is commonly attributed to the existence of various institutional factors and rigidities, which prevent wages from falling. The most often cited factor is probably the influence of South Africa's strong trade unions. Indeed, Schultz & Mwabu (1998:680) estimated that "among male African workers in the bottom decile of the wage distribution, union membership was associated (in 1993) with wages that were 145% higher than those of comparable non-union workers". This is likely to play a part in the existence of high urban unemployment, as are other structural rigidities arising from such factors as a lack of skills and basic education. Perhaps more fundamentally, lack of labour demand is also a feature of urban labour markets, or at least lack of labour demand in skill areas appropriate to the unemployed.

At first sight it is more difficult to explain the non-clearing of the rural market for labour. In general the rural labour market is characterised by a much narrower skill base among the work-force than exists in urban areas, and consequently there is less scope for structural differences to occur between labour demand and supply. Furthermore, there is also an apparant lack of enforced regulation in the rural labour market, which means that wages should not be prevented from falling in the same way that they might be in urban markets [Klasen & Woolard (1998:4)]. A number of issues have emerged in the debate however, which seem important in explaining the rural phenomenon.

Following from the above observation on coping strategies, it is probable that a proportion of rural unemployment may in fact be disguised urban unemployment, whereby those unable to find jobs in urban areas have moved to rural households (or remained there) to take advantage of support mechanisms.[15] Furthermore, by nature these individuals are likely to have very low job prospects, and in a sense this could mean a sort of self-selection of the least able into rural areas. Meanwhile, there may be a corresponding migration of the most able in the other direction, to search for more rewarding

[14] In particular this adapting is likely to take place through the delaying of marriage and independent household formation by the unemployed, and consequent attachment for longer periods to the household of their parents or other relatives [Wittenburg (1999), Klasen & Woolard (1998)].
[15] See Klasen & Woolard (1998:23-29) for a detailed discussion of this argument.

employment or to undertake further education and/or training. Consequently, those left in the rural labour market are likely to be the least skilled and least flexible of workers, and to be characterised by a high level of discouragement. If this is the case then it would certainly hinder the flexibility of the rural labour market, and may go some way towards explaining the high observed unemployment.

This explanation is unlikely to account for the whole rural unemployment problem, and figures on agricultural employment indicate that rural labour demand has also been falling over quite a substantial period of time, in part due to technological factors.[16] Serving to exacerbate the effect of falling demand, Klasen & Woolard (1998:4-5) also suggest the persistence of rural labour market rigidities that were generated by "*apartheid* residential policies that created extreme labour market segmentation between the labour markets of the former homelands and the labour markets of the non-homeland rural areas".

2.4.3 The Racial Divide

Perhaps the most significant feature of the South African unemployment figures is the degree to which they diverge by race. It is indeed clear that race is still a fundamental determinant of employment in South Africa. This may be related in part to aspects of location, and the fact that larger numbers of the African population live in rural areas (and particularly the impoverished former homeland areas) than do other population groups. Indeed, the two provinces hardest hit by unemployment are noticeably those that incorporate much of the former homeland areas, which were subject to enforced overcrowding of African labour during the apartheid era.

While location may play a part, the potential for migration should ameliorate this impact on unemployment between races. Indeed, the explanation is likely to lie more with the skills and educational characteristics of the different race groups, and the persistence through these characteristics of the disadvantages enforced on the African population (and to a lesser extent the coloured and Asian populations) during apartheid. Average completed years of education are just 7.6 and 7.7 for the African and coloured populations

respectively, as opposed to 10.4 for Asians, and 12.1 for whites. Where a particular population group has such an educational disadvantage, this is bound to manifest itself in higher unemployment.

Such a disparity will take time to address, and in this regard Deaton & Case (1998:24) argue that still today educational resources are "sharply different by race, with pupil-teacher ratios in Black schools more than twice as high as those in White schools". Moreover, in this environment they find that "high pupil-teacher ratios discourage education attainment conditional on age, lower test scores, lower the probability of being enrolled in education, and discourage parents from making complementary expenditures on children's education" [*Ibid.*]. From table 2.4 we see that the unemployment rates in South Africa do not fall consistently with additional years of education, and this also supports the notion that quality of education is important. It is generally acknowledged that a key element in tackling unemployment, and in particular in lowering it's incidence among certain population groups, must therefore lie with policies to improve and equalise educational opportunity and quality.

2.4.4 Issues of Gender and Age

Finally, two further important issues that have stimulated much debate are the gender and age profiles of the South African unemployed. The figures show that unemployment disproportionately effects females, and even more striking is the significant concentration of unemployment among young South Africans. The gender differential is certainly a concern, particularly in the light of a growing recognition of the importance to the development process of employment opportunities for women. Moreover, the causes of the differential are difficult to determine, and there is uncertainty around the degree to which these figures reflect gender discrimination in the labour market.

A significant and interesting study has been carried out in this regard by Winter (1998), where OHS 1994 data were used to study gender differences in formal labour market participation and pay. This study highlights the alarmingly low participation rates of

[16] For a more detailed discussion of the factors that may have contributed to this falling labour demand, see

African females, although it finds that the wage gap between African men and women is essentially insignificant [Winter (1998:31)].[17] This suggests that African women have a higher reservation wage than men, a fact that can be reconciled with issues of childcare and domestic duties. Differences in unemployment rates, however, are not analysed by Winter, and there is certainly a possibility that discrimination occurs in the selection of employees. However, the relative skill characteristics of the two genders and their correspondance with labour demand is also likely to play an important role.

The severity of unemployment among the youth of the population is such that it has generated "the perception in many quarters that the unemployment challenge is overwhelmingly a youth unemployment problem" [ILO (1996:117)]. Indeed, while the problem is again most vividly noticeable among young Africans, the figures suggest that unemployment disproportionately effects the young of all population groups. There have been various arguments as to the cause of the youth unemployment phenomenon, mostly based around the common perception of a generally inflexible labour market. One such argument is that the labour market is not flexible enough to lower wages among young inexperienced workers, and consequently the youth are paid higher 'adult' wages, which creates unemployment [ILO (1996:117)]. Another suggestion in a similar vein is that young people emerging from school typically have high reservation wages, which further impede downward flexibility in wages among the young [*Ibid.*].

Inflexible labour markets have certainly been a focus of much of the debate on unemployment in general, and indeed it is likely to be the case that rigidities which prevent wages and employment from adjusting will particularly harm younger members of the labour force. Older workers have the advantage of being more established in the workforce, and where trade unions have a strong influence this can be an important factor in placing new-entrants the bottom of the queue. Furthermore, when education and skills levels are generally low, the young may have little way of pushing themselves in front of older workers, because experience then becomes an even greater advantage.

Klasen & Woolard (1998:42-45).

[17]The absense of a gender wage gap is however partly attributed to the educational advantage that African women show over men in the sample. For Whites, who have the same average educational achievement among males and females the wage gap is indeed large, and is suggestive of a good degree of gender discrimination [Winter (1998:31)].

This reasoning might imply that one way to tackle unemployment (and youth unemployment in particular) is to implement policies that will enhance the flexibility of markets and allow wages to fall. While falling wages among the young may ease the problem and allow the market to absorb some of the pool of unemployed, there are inherent dangers with such policies in that they may hinder the goals of poverty and inequality reduction. This is especially so if falling wages at the bottom end of the labour market are transmitted to older workers with families (and possibly unemployed family members) to support, and already living close to subsistence. Indeed, to be sensitive to this concern, there have even been recommendations (from the traditionally market-oriented instiutions of the World Bank and IMF) that age-based wage differentials should be introduced to enable wages among the young to fall relative to older workers [ILO (1996:117)]. More generally, many argue that there must be a superior solution for creating jobs than a simple reliance on flexible markets, and one which is sensitive both to the need to reduce unemployment, and to the importance of not exacerbating poverty and inequality.

In summary, the debate surrounding South African unemployment has been stimulated both by the high unemployment figures themselves, and by as a number of important and noticeable features of the incidence of unemployment. Unsurprisingly the debate has therefore been wide and varied, although as with most unemployment debates it has focussed on motivation and the voluntary nature of unemployment, the inflexibility of labour markets, and various structural issues and disparities. No clear consensus emerges on these issues or the correct approach to deal with them, and this perhaps emphasises the need for a continued and close analysis of the problem, so that policies can be developed which will effectively tackle unemployment without unnecessarily exacerbating other (and related) important societal concerns. It is with various elements of this debate in mind that I move to investigate the process of labour supply among the African population group in South Africa, starting in the next chapter with a theoretical introduction to the modelling of the labour supply decision.

Chapter Three

Labour Supply Theory and Modelling

3.1 Introduction

The subject of labour supply has always stimulated great interest among economists, and one of the consequences of this has been the development of a huge theoretical and empirical literature on labour supply issues. Besides the acknowledged importance of labour supply as a research area, a key reason for the evolution of such a large and wide body of literature are the inherent difficulties that surround this subject. Modelling the decisions that individuals make with regard to their supply of labour is indeed a tricky process, with many complications to consider. Consequently much of the literature on labour supply has been devoted to addressing some of these difficulties, and this has resulted in the refinement of various econometric techniques. It is worthwhile therefore, to make a careful consideration of the underlying theory, and how this relates to the various empirical specifications of labour supply models that have evolved.

In this chapter I start by outlining basic neo-classical theory on the supply of labour, as this topic provides the optimisation framework in which most of the empirical work is grounded. I then examine how we can seek to extend this incomplete theoretical model, to make it relatively compatible with the empirical realities of labour markets. I discuss some of the empirical issues that arise when we attempt to do this, and arrive at a series of different approaches that are conventionally used for modelling labour supply. Finally, I suggest that these neo-classical models may be largely unsuitable for modelling the supply decision in a country such as South Africa, where the labour market is characterised by a high degree of rationing. I therefore suggest the careful use of such models, and argue for a more descriptive analysis, which uses the conventional techniques in a way more suitable to the South African labour market.

3.2 Neo-classical Labour Supply Theory

The static neo-classical model of labour supply is the common starting point for most analyses of the labour supply decision, and is rooted in the neo-classical framework of optimisation. It is assumed that each individual implicitly decides how many hours of labour to supply through a simple maximisation of his/her utility function. Essentially the decision is between consumption and leisure, as reflected in the standard utility function:

$$U = U(C,L) \tag{3.1}$$

Here, C represents a composite bundle of consumption goods consumed in each period, and L is the amount of leisure (in hours) undertaken in each period. Each individual will aim to maximise his/her utility derived from consumption and leisure, and as described by this function. Leisure is thus assumed to hold the properties of a normal good, in that utility rises as more is consumed, *ceteris paribus*. There is an inherent trade-off between C and L in this relationship, however, which reflects the fact that to earn money requires a foregoing of leisure to spend time in the labour market.[1] This trade-off can be represented in the dual constraint:

$$PC = WH + V \tag{3.2}$$

$$T \equiv H + L \tag{3.3}$$

In equation 3.2, P is the general price level, W the wage rate, H the number of hours of labour supplied, and V the amount of non-labour income that the individual receives. This constraint simply states, therefore, that the cost of consumption (PC) must equal the total income of the individual; where this income consists of a combination of labour market earnings (WH), and non-labour income (V). In the identity 3.3, T is the time allocation of

[1] This implies that work must be seen as an economic 'bad', as it denies the consumption of the normal good leisure. This assumption in itself is open to scrutiny, as for many people work may bring positive benefits. Dilnot & Duncan (1993) make this point from the perspective of the economic psychology literature.

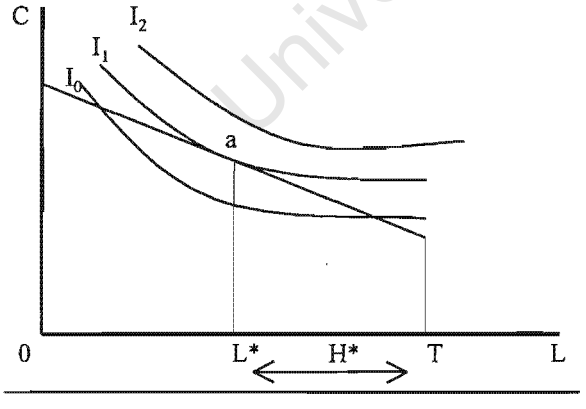
each individual[2], and we assume that it is split exclusively between work (H) and leisure (L).[3]

A solution to the maximisation of the utility function, subject to the income constraint, can be shown to occur when the marginal rate of substitution between consumption and leisure is equal to the real wage rate:[4]

$$M_{C,L} = (dU/dL)/(dU/dC) = W/P \qquad (3.4)$$

This is equivalent to a point of tangency between the individual's budget constraint, and his/her indifference curve, as seen at point *a* in figure 3.1. The slope of the budget constraint is the real wage rate (W/P), and the marginal rate of substitution between consumption and leisure ($M_{C,L}$) is simply the slope of the indifference curve. The individual will choose a combination of leisure and consumption which leaves him/her on the highest possible indifference curve, and we can see that at this tangency $M_{C,L} = W/P$.[5] Figure 3.1 also shows us the desired labour supply of the individual (H*), given the optimal choice of leisure and consumption.

Figure 3.1 Determination of Desired Hours



[2] This time allocation reflects the natural constraint imposed by there being only 24 hours in each day, seven days in each week etc..

[3] A potential variation on this assumption (which is particularly relevant for models of married women's labour supply) is to instead make a threefold division of time; into labour market hours, hours of work in the home, and true leisure [Polachek & Siebert (1993:108-112)].

[4] Here, $M_{C,L}$ refers to the marginal rate of substitution between consumption and leisure.

If we now combine equations 3.2 and 3.4 with the time identity 3.3, it is possible to derive expressions for the individual's optimal quantities of consumption and labour hours in terms of the other variables involved. Thus we can say that $H^* = H(P,W,V)$. Furthermore, by defining $w = W/P$ as the real wage rate and $v = V/P$ as real non-labour income, we can simplify further and state that:

$$H^* = H(w,v) \qquad (3.5)$$

In 3.5 we have derived a function for desired labour supply in terms of real wages and non-labour income. Thus in a neo-classical optimisation framework, we can see that the hours of labour supplied are theoretically determined by the wage rate offered, and the non-labour income of each individual. Although this result may seem overly simple, its strength lies in its flexibility, and the fact that it can be modified in many ways to incorporate different facets of the labour supply decision. As such, it simply provides a backbone for our thinking about the labour supply decision, and allows us to relate further modifications back to a solid theoretical framework.

3.3 From Theory to Empirical Modelling

The static neo-classical model outlined above provides a foundation for studying the labour supply decision. However, with respect to modelling what we actually observe in labour markets it is clearly incomplete. Thus we need to make certain alterations to the basic model, to be consistent with an analysis of the empirical realities of the labour market. While there are many points that can be considered here, there are two that stand out above the others and these will be discussed in this section.

[5] Note again that this optimisation is dependent on behavioural assumptions determining the shape of the indifference curve. In particular, the indifference curve is assumed to be convex to the origin, which implies that leisure is a normal good and thus work an economic 'bad'.

3.3.1 Specifying the Model to Account for Different Individuals

An immediate problem with applying this framework to an empirical modelling of the labour market is that the theory is expressed in terms of a given individual, and as such it treats all individuals identically. In reality the labour market consists of different individuals with inherently different tastes and propensities to work. Thus we would expect to find that two people with the same non-labour income, and offered the same wage rate, would actually desire to work for different amounts of their time. With reference to figure 3.1, we can think of two individuals facing the same budget constraint, but with different utility functions and therefore a different set of indifference curves. By treating all individuals equally, the simple theoretical model does not account for these differences in individual characteristics, and consequently equation 3.5 cannot provide a full model of labour supply behaviour. An empirical model must therefore make certain adjustments, so that it can cope with differences among individuals.

Some of the differences between individuals can be linked to various observable characteristics of individuals, and so this problem can be addressed by including these characteristics as additional explanatory variables on the right-hand-side of the model [Killingsworth (1983:75)].[6] For example, age, sex and race are characteristics that may be associated with differing propensities and tastes for supplying labour. For female labour supply in particular (although also for male labour supply), additional important factors may be whether the individual is married, and whether she has children. If these factors are included on the right-hand-side of the model, then the results will reflect the influence that certain observable characteristics have on labour supply, and in this sense allow for differing utility functions among people with different characteristics.

The problem is not solved completely, however, as there will still be differences between individuals caused by unobservable characteristics. Tastes and preferences for work are not something that can be measured specifically, and differences in these across individuals are likely to stem from unobservable characteristics in addition to the observable ones mentioned above. In an empirical modelling of the labour supply

decision, the differences between individuals attributable to these unmeasured factors are represented by a random error term with mean of zero.

Making these two adjustments, equation 3.5 can be re-specified as:

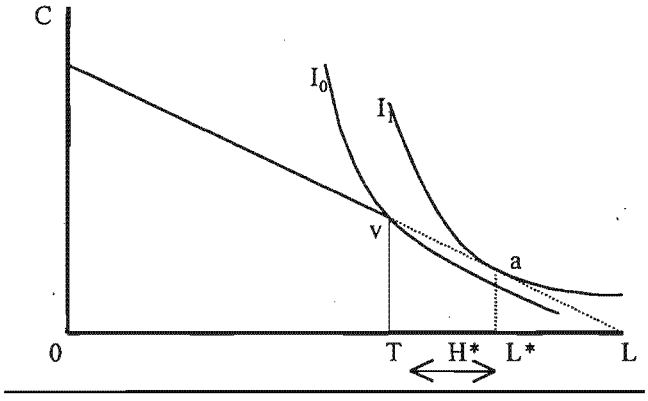
$$H^* = H(w,v,\mathbf{Z},e) \qquad (3.6)$$

Here, **Z** is a vector of observable characteristics that are believed to influence the labour supply decision, and e is the random error term to account for unobservable differences between individuals.

3.3.2 Addressing the Non-Participation Problem

In real labour markets we generally observe a situation where only a proportion of individuals actually participate in the labour market. This presents a second problem, in that the simple theoretical model does not explicitly consider corner solutions to the individual's maximisation problem, and thus assumes that every individual will work. The model says nothing, for example, about a situation in which an individual's demand for leisure is unsatisfied even when allocating all available time to leisure [Killingsworth (1983:73)].

Figure 3.2 Negative Desired Hours



[6] Alternatively, we could focus a given study specifically on individuals who are alike in certain observable

Such a situation can be seen to occur for an individual with the set of indifference curves in figure 3.2. Here the tangency between the individual's $M_{C,L}$ and the real wage rate occurs at point *a*, where desired labour supply (H*) would actually be negative. The optimal obtainable point would be at point *v*, with zero hours of labour supplied. Although a non-tangency, point *v* is where we would observe this individual in practice, and this suggests that we need to consider an additional refinement to explicitly account for such individuals in the empirical model.[7]

A relatively straightforward way of incorporating this participation decision is to specify the model such that actual labour supply will equal desired labour if desired labour supply is greater than zero, and will equal zero if desired labour supply is less than or equal to zero:

$$H = H^* = H(w,v,\mathbf{Z},e) \quad \text{if} \quad H(w,v,\mathbf{Z},e) > 0 \qquad (3.7)$$

$$H = 0 \quad \text{if} \quad H(w,v,\mathbf{Z},e) \leq 0 \qquad (3.8)$$

Theoretically, we can think of this dichotomy between participants and non-participants in terms of a reservation wage, whereby an individual will only desire to work if the offered wage exceeds a reservation wage specifically applicable to that individual. In an optimisation framework, individuals who choose not to work presumably do so because the marginal value of their time is greater than the real wage rate that is offered.[8] Thus, each individual's reservation wage will equal the value of his/her marginal rate of substitution between consumption and leisure when all of his/her time is allocated to leisure. This refinement to the model allows us to consider cases where an individual chooses not to supply any labour, and thus gives us a more complete model of labour supply.

---

characteristics; for example white married women in urban areas.

[7] This also highlights a further problem, in that it is only possible to observe actual labour supply, and not the desired labour supply to which the theoretical model actually refers. I will return to this point in a broader sense in the next section.

3.4 The Practicalities of Modelling Labour Supply

In equations 3.7 and 3.8 we have a model of labour supply which is grounded in neo-classical optimisation theory, and reflects both the inherent differences between individuals in the labour market, and the reality that some people will desire not to participate. It can therefore be seen as a fairly powerful model, and indeed it is this general model on which the majority of econometric studies of labour supply are based. Estimation of such a model is not straightforward, however, and the exact design of the model to counter various data and specification bias problems has stimulated a great deal of debate in the literature.

Indeed, problems of measurement, sample selection and functional form tend to be intimately related when modelling labour supply, and consequently "these problems had to be confronted using not only a unified and sophisticated theoretical methodology, but also improved econometric techniques specifically designed to suit that methodology" [Killingsworth (1983:68)]. I will briefly discuss some of these issues in this section, leading to an outline of the conventional ways of estimating labour supply functions in the light of such problems.

The difficulty in modelling labour supply arises largely from the fact that an individual's labour supply is comprised of two separate, but related, decisions. Firstly there is the decision to participate in the labour force, and conditional on this is the decision of how many labour hours to supply. In particular, we have to decide how best to model this dual decision given the likelihood that participation and hours are determined (at least partly) simultaneously. The distinction between the two was largely ignored in early studies of labour supply, although in practice it is absolutely fundamental to correct estimation. In particular, the distinction highlights the problem of missing wages for non-workers, and has stimulated the need to develop techniques to overcome the selection bias that this presents.[9]

---

[8] In other words, the individual's indifference curve is either tangent to the budget constraint at the point where all time is allocated to leisure (point v in figure 3.2), or would be at some point after this (e.g. point a in figure 3.2).

[9] Indeed, Heckman (1993:117) argues that this distinction has elevated the consequences and causes of self-selection bias in estimating wage and labour supply functions to a central research problem in economics.

3.4.1 Sample Selection Bias

If the labour supply sample is restricted only to workers, then simple regression estimates will suffer from selectivity bias, deriving from the fact that the error term cannot be considered of mean zero within that sample. By their very nature participants are likely to have higher tastes for work (and other unobservable characteristics) than non-participants, and thus the average value of the error term will tend to be positive if non-participants are discounted. In line with the theoretical model, this amounts to selection with respect to the reservation wage of individuals, with non-participants likely to have higher reservation wages on average than participants. When modelling the decision of how many hours of labour to supply, we must therefore ensure that we account for the process by which individuals are selected into the labour market in the first place (i.e. the participation decision).

There is however a related problem with respect to modelling the participation decision, in that participation is itself dependent on the offered wage, which is unobserved for non-participants. This can be addressed by incorporating a standard wage equation into the analysis, to develop an imputed wage for each individual that reflects their specific characteristics.[10] However, such an equation will again suffer from selectivity bias, because those who find themselves in the workforce are generally likely to differ in certain unobservable characteristics (such as motivation or competence) to non-workers. Consequently the error term will again be non-zero for the sample of workers, and we must therefore incorporate techniques to correct for selection bias at this stage.

Selection bias is certainly a significant issue when performing any estimations using observed hours or wages. Indeed, given the fact that non-participation characterises all labour markets, Heckman (1993:118) argues that "one cannot have an economically meaningful model of labour supply in which there is no selection on either wages or reservation wages".

[10] In other words, a standard Mincerian human capital function.

3.4.2 Dealing with a Censored Dependant Variable

The very nature of this dual decision provides an immediate problem of missing data for those who choose not to participate, and this issue must also be addressed. With respect to the dependant variable, we have already noted a dichotomy for non-participants between their desired labour supply, and the hours of labour that are actually observed. This is resolved in the model by making a 'cut-off' assumption, whereby those who desire to supply zero or negative hours of work are represented by zero hours (the best data available). We therefore have a dependent variable which is censored (in this case truncated from below), and this requires special techniques of estimation, as "conventional regression methods fail to account for the qualitative difference between *limit* (zero) observations and *nonlimit* (continuous) observations" [Greene (1990:725)].

Following Breen (1996), there are essentially two groups of models that can be used for estimation where the dependent variable is censored. The first of these are called 'Tobit' models, and are commonly used where there is a simple censoring of the latent dependent variable at a specific value.[11] These models produce results whereby a single set of variables and coefficients "determine both the probability of truncation and the expected value of the realised dependent variable, conditional on its having been observed" [Breen (1996:33)]. A second set of models, known as 'generalised Tobit' or 'sample selection' models, relax this constraint and allow the effects of variables to be different at each stage. These models are therefore more consistent with the view that the two decisions show a degree of independence, and in particular with the idea that those in the labour force are there through a process of self-selection. Which model we use really depends therefore on how we view the labour supply decision and, in particular, the interdependence between the participation and hours decisions.

[11] The Tobit model is also sometimes referred to as Tobin's Probit, and relates to Tobin's (1958) study in which he applied this model to a modelling of the ratio of disposable income spent on durable goods. In this case, consumer durable expenditure as a dependant variable was truncated from below in the sense that for certain households expenditure was observed as zero [Breen (1996:6)].

3.4.3 Tobit Models of Labour Supply

The simple Tobit model of labour supply follows directly from the specification of equations 3.7 and 3.8:[12]

$$H_i^* = \mathbf{Z}_i'\beta + e_i \qquad (3.9)$$

Where; $H_i = H_i^*$ if $H_i^* > 0$

$H_i = 0$ if $H_i^* \leq 0$

Here, desired hours (H*) can be seen as a latent variable, and observed labour hours are truncated in the sense that they are zero when H* is less than or equal to zero.

This can be estimated through a two-stage procedure (commonly referred to as the Heckman two-step procedure), whereby a participation equation is first estimated, and the resulting probability estimates used to derive estimates of the inverse Mills ratio (λ) for observations with positive hours.[13] A labour supply equation can then be estimated by OLS for all observations with non-zero hours, with $\hat{\lambda}$ included on the right hand side to correct for the probability of selection. Unfortunately this method gives inconsistent estimates of the standard errors, and while it is possible to correct these errors, it is more common to instead estimate the two equations simultaneously using maximum likelihood [Breen (1996)].

Maximum likelihood estimation of the Tobit model makes simultaneous use of information on participation and hours, and produces one set of parameters from which both decisions can be analysed. Indeed, both methods of estimating the simple tobit model imply a strong association between the participation decision and the labour hours decision. Since the model is theoretically derived from a comparison of reservation and

[12] For ease of notation, the vector **Z** now includes all of the parameters thought to determine labour supply, including the wage rate (w) and non-labour income (v).

[13] The inverse Mills ratio is the ratio of the probability density function to the probability distribution function, and essentially incorporates information on the probability of selection into the labour supply equation. For a detailed discussion of the statistical theory behind the Heckman two-step procedure see Breen (1996:12-17).

market wages, the two decisions are inextricably linked, and the structure of the model is such that the parameters in the hours equation will be proportional to the parameters in the implicit participation equation [Zabel (1993)].[14]

An association between the two might seem reasonable, given that the two decisions are certainly related in some way. However, there are strong arguments that suggest a degree of independence between the decisions, the most significant being the existence of fixed costs to working, and the fact that there may be constraints on the number of hours that individuals can choose to work [Zabel (1993)]. It is therefore prudent to also investigate models that relax the strict proportionality assumption imposed on the parameters by Tobit models.[15]

3.4.4 Sample Selection Models of Labour Supply

In sample selection (or generalised Tobit) models of labour supply the situation is allowed to be more complex. In a sense it moves away from a simple censoring of the dependent variable, and introduces a censoring which is itself dependent on a group of variables, which may or may not be the same as those that are deemed to determine the hours supplied. Thus it allows both the explanatory variables and the estimated parameters to differ across the two equations. The decisions are still related, however, through the disturbance terms in the two equations [Zabel (1993)].

Following Breen (1996), the model has two stages. It is structured around a dummy variable for participation (P), which is equal to one when a latent variable P* (as determined by variables **X**) is greater than zero.[16] The hours variable (H) is then only

[14] For a worked example which illustrates this relationship, see Breen (1996:24-27).

[15] Indeed, many empirical studies of labour supply focus on more than one type of model, which is a reflection of the uncertainty and lack of consensus on this issue. Zabel (1993) has investigated the relationship between the participation and hours decisions for four different models of labour supply behaviour (including 'fixed-cost' and 'hours-constraint' models), each displaying different restrictions on the two sets of parameters. He found the participation and hours decisions not to be as strongly tied as the Tobit model specifies, and suggests that the generalised Tobit model, which has no restrictions, is therefore a useful one to estimate.

[16] The latent variable P* can be thought of as some measure of the propensity of each individual to work. In this sense it is loosely related to the reservation wage concept discussed earlier.

observed when the participation dummy is positive, and is itself equal to another latent variable H* (as determined by variables **Z**).[17] Formally, it can be expressed as:

$$P_i^* = \mathbf{X}_i'\alpha + u_i \qquad (3.10)$$

Where; $P_i = 0$ if $P_i^* \leq 0$

$P_i = 1$ if $P_i^* > 0$

And;

$$H_i^* = \mathbf{Z}_i'\beta + e_i \qquad (3.11)$$

Where; $H_i = H_i^*$ if $P_i = 1$

H_i not observed if $P_i = 0$

Both of the error terms in this model are assumed to be independently and normally distributed, with mean of zero and constant variance. Furthermore, the two error terms are assumed to have correlation ρ, which is the mechanism through which the two decisions are related, and reflects the fact that both decisions are effected by similar unobservable factors [Breen (1996)].[18]

To estimate this model, we can again use the Heckman two-step procedure, although it is slightly more complicated than when estimating the simple Tobit, due to the relationship between the two error terms. As with the simple Tobit model, however, the standard errors will be incorrect, and so it is necessary to make some alterations. Alternatively, we can estimate this sample selection equation simultaneously by full information maximum likelihood (FIML), and given that the FIML estimates will be asymptotically unbiased and efficient, this method is generally preferred [Breen (1996:42), Maglad (1998:11)].[19]

[17] The latent variable H*, as before, represents the desired labour hours of each individual.
[18] Strictly, therefore, the error terms are assumed to be jointly normally distributed.
[19] For a discussion of technicalities behind both methods of estimation, see Breen (1996:36-42)

3.5 Labour Market Rationing and the Need to Tailor Conventional Models

As is clear from the above discussion, an intricate web of problems and issues complicates the process of modelling labour supply. Indeed, it is the above difficulties that have made the study of labour supply such a rich one, and after decades of discussion there is certainly still no easy and universally accepted solution. I have identified two commonly used methodologies for estimating labour supply functions, and my analysis of South African labour supply will focus on these techniques. Their criticisms and difficulties should not be ignored, however, and any analysis must therefore be extremely careful in how it uses these models and draws conclusions from them.

3.5.1 Criticisms of the Neo-classical Model

There are many criticisms of such a neo-classical modelling of the labour supply, some of which I have already touched upon. Dilnot and Duncan (1993) challenge some of the assumptions on which the standard model is based, and are particularly concerned that it is founded on a purely financial notion of the benefits from working. A more reasonable model, they argue, "would acknowledge that the rewards for working can be expressed as much in terms of social and personal esteem as on purely financial grounds" [Dilnot & Duncan (1993:689)].[20] They also bring attention to the assumption that individuals are sufficiently free to realise their desired labour supply, an observation that can encompass a number of different circumstances, grouped under the heading of 'labour-market rationing'.[21]

Indeed, in the South African context the occurrence of labour market rationing constitutes the most important criticism of the neo-classical model. Broadly speaking, rationing in the labour market can manifest itself in two ways. The first is through constraints imposed

[20] The idea that the rewards to work are more than simply financial is supported by findings on the relationship between unemployment and general well-being, a popular topic in the economic psychology literature. For example, Clarke & Oswald (1994:658) find that "unemployed people in Great Britain in 1991 have much lower levels of mental well-being than those in work".

[21] Killingsworth (1983:45-66) provides a detailed discussion of the influence of different forms of rationing and other non-tangency phenomena within the neo-classical framework.

by employers, institutions, and society on the number of hours that individuals can choose to work. The second is through the phenomenon of unemployment, and the inability of individuals to find suitable work. An implicit assumption of the standard model is that each individual faces a continuous and 'well-behaved' budget line, on which he/she has full information and can make utility-maximising decisions accordingly. In reality this is unlikely to be the case, as both job opportunities and choice over number of hours tend to be constrained.

The first form of rationing is best described as a 'take-it-or leave-it' regime, whereby the option of working 30 hours a week rather than 40, for example, is simply not exercisable. There seems to be general agreement, however, that this form of rationing doesn't pose very serious problems to the modelling framework. Indeed, Killingsworth (1983) argues that this form of rationing can be thought of as an equilibrium outcome, in the sense that it may well be the result of voluntary choice, despite the seeming constraint. The argument is that if large numbers of individuals were rationed in this way, then they would have an incentive to seek better solutions, and firms an incentive to supply them. Consequently, non-tangencies resulting from this form of rationing must be seen as equilibria if they persist in the longer term. I am cautious to accept this argument totally, however, as it is a very neo-classical response to the problem, and doesn't account for the power relationships inherent in the workplace between employers and employees. We should therefore at least be aware of the possibility of constraints on choice of hours, and in particular consider the likelihood that individuals have more immediate choice over their participation than they do over hours worked.

3.5.2 Unemployment as a Rationing Problem

Given the subject of this study, my interest really lies with the second form of rationing, that of unemployment. When individuals who are seeking work cannot find suitable employment, then we have the concept of involuntary unemployment. This concept, however, is alien to the neo-classical model of the labour supply, and indeed to much of neo-classical economics in general. As we have seen in the above discussion, individuals

are assumed either to participate or not-participate in the labour market. Those that do participate then supply a certain number if hours, and are paid an observed wage. There is no scope, however, for those who choose to participate, and yet in the face of rationing cannot find work. These individuals should be classed as participants, but the fact that their hours and wage are not observed results in them being counted along with true non-participants.

In certain economies where the numbers of involuntarily unemployed are low, this rationing does not present a huge problem in modelling labour supply. Indeed, if the aim is simply to investigate the determinants of hours supplied for those with jobs, conditional on them having jobs, then the models described above can still be effective even in economies with high unemployment. However, the aim of this study is to examine the unemployment problem in South Africa from the perspective of the supply-side of the labour market. It is therefore important to be able to separate non-participants, from unemployed participants, from those who are employed, and to investigate the factors that are associated with each stage. In an economy such as South Africa, where the dichotomy between participation and employment is so large, the conventional neo-classical models outlined above are therefore clearly insufficient to describe the full process of supplying labour.

In response to this deficiency, Andersson (1993) suggests a particular generalisation of the Tobit model that was developed by Cragg (1971). Known as the Double- Hurdle model, this adds an extra condition to the model presented in equations 3.10 and 3.11, and consequently allows participants who are seeking work but cannot find it to be distinguished from true non-participants.[22] Simultaneous estimation of this model in its entirety turns out to be complicated, but my analysis will attempt to accomplish a similar aim in a more general way, by separating out each of the steps.

In particular, I will pay careful attention to describing the South African labour market realities, and use the techniques described above to investigate each stage of the labour supply process. This should allow me to generate a picture of each of the three

[22] This model is presented in *Appendix 2*.

components of labour supply in the market; the desire to find work, the ability to find work, and the amount of hours that are ultimately supplied. The techniques outlined in this section can certainly help to do this, but we must be careful to relate them properly to the particular problems of the labour market in question. As noted by Duncan and Dilnot (1993:708), "the key is to choose methods and econometric techniques appropriate to the behaviour of the demographic group under discussion." In this case therefore, we need to be aware of the large degree of rationing observed amongst African workers in the South African economy, and to ensure that our methodologies and interpretation of results reflects this.

Chapter Four

An Introductory Analysis of the Data

4.1 Introduction

Before undertaking an econometric estimation of the labour supply process using the modelling techniques outlined in the previous chapter, it is important to describe the data from which I will estimate my models. I start in this chapter with a brief introduction to the OHS 1995 data set and the survey on which it is based, incorporating a more general discussion of some of the sampling issues that arise when analysing survey data. In chapter two I have already used this data to present the description of unemployment. After introducing the OHS data, I describe the customisation of the data set and generation of variables relevant for labour supply analysis, before moving on to present some summary statistics of these variables for the African population. This analysis is split by male and female, and the focus is on certain key variables that might be considered to determine wages, participation, the ability to find a job, and the hours of labour that are ultimately supplied. Finally, I briefly introduce a few points of awareness concerning the potential for measurement error in some of the labour supply variables.

4.2 The Data Set

4.2.1 The OHS 1995 Survey

The data that I will use for my labour supply analysis is from the 1995 October Household Survey, which was conducted by the CSS. The OHS is an annual survey that was first carried out in 1993, when one of its main aims was to collect information on the size of the informal sector [Klasen & Woolard (1999:8)]. The survey has evolved slightly in the following years, and the questionnaire now encompasses a wide range of issues concerning employment, living standards, and general household and individual

characteristics. Unfortunately the data from the 1996 and 1997 surveys has not yet been made available for analysis, and so the most recent data accessible is from the 1995 OHS.

The 1995 survey questionnaire is split into five sections. The first section is at a household level, and deals with the nature of the dwelling and the basic services and facilities available. It also asks questions to ascertain the perceived quality of life of household members. The second section is then aimed at collecting information on the personal characteristics of each individual in the household, including issues such as family relationships, health and education. The third section is the most important for our purposes, as it asks detailed questions on labour market activities, and gathers information pertaining to workers, the unemployed, and the economically inactive. Finally, sections four and five deal specifically with deaths and births in the household over the last year. The 1995 OHS also included an Income and Expenditure Survey (IES), which incorporated 98% of households sampled in the main survey, and asked detailed questions on household income and expenditures.

4.2.2 Sampling Issues

The sample itself covered 30,000 households, and included just over 100,000 individuals. The survey was conducted according to a two-stage sample design, with ten households selected from each of 3,000 clusters. This method of sampling is common practice in such surveys, as it allows for increased efficiency in interviewing, and also the possibility of collecting community level information in individual clusters. Clustering of this nature will however generate groups of observations that cannot be considered wholly independent, as observations from the same cluster will tend to show a degree of homogeneity [Deaton (1997)]. There will therefore be an overstatement of independent observations in the sample, and as a consequence standard formulas will inherently underestimate the variance of parameters. When using survey data of this nature we must recognise this limitation, and where possible make the necessary adjustments.[1]

[1] See Deaton (1997) for a comprehensive discussion of survey design and estimation. The computer package (STATA) that I will be using makes it possible to adjust standard errors for clustering with simple OLS and probit regressions, and with the computation of basic summary statistics. The more complicated Tobit and sample selection regressions, however, do not include this option. While this will not effect the

Also included in the OHS data set are a series of sample weights, which scale the sample up to represent the full population, and correct for over and under-representation of certain types of households. Each observation has an attached weight, which corresponds to the proportion of the total population that that observation represents. There are many issues to consider when deciding whether to incorporate weights into our analysis, including the possibility that the weights themselves may be inaccurate.[2] As general guidance, however, Deaton (1997:71) recommends the use of weights if the purpose of the analysis is descriptive. This is very much the case, and I will therefore adjust all of my statistics and regressions (using the OHS weights) to be representative at the level of the population.[3]

4.2.3 Identification of Labour Supply Variables

There are naturally certain variables which are of interest when conducting an analysis of labour supply, and these need to be identified so that the data set can be customised to our purposes. First of all it is necessary to have variables for participation, employment, and hours supplied. These correspond to the three decision stages of the labour supply process, and as such will constitute the dependent variables in the analysis. We then need to ask which factors are likely to impart an influence on the respective decisions, and from these we will have a set of independent variables.

We know from the discussion of the previous chapter that neo-classical theory identifies the real wage rate and the level of real non-labour income as the key determinants of labour supply (equation 3.5). In terms of the participation and hours decisions these will therefore be central, although given the cross-sectional nature of the data a weighting by the price level is not necessary.[4] We also noted in chapter three that there are likely to be certain observable characteristics which will influence the propensity of individuals to supply labour, and which should therefore additionally be included on the right-hand side

parameters as such, it will influence the standard errors, and the precision of estimates and significance tests from these models should therefore be evaluated with this in mind.

[2] Again, see Deaton (1997) for a full and technical discussion of such issues.

[3] It is again possible to adjust parameters and standard errors correctly for weighting when estimating OLS and probit regressions using STATA, or when producing summary statistics. In the more complicated models, however, an adjustment for weighting will correctly alter the parameters, but simply re-scale the standard errors for an increased number of observations.

(equation 3.6). Entirely separate analyses will be carried out for males and females to allow for differences between these two groups, but factors of age, marriage, children and urban/rural location will be included as additional independent variables in the participation and hours equations. In terms of the second stage of the labour supply process the variables will differ slightly. This stage pertains to the ability of participants to find employment, and hence factors of importance will be education, experience, and location (provincial and urban/rural).

4.2.4 Customising the Data Set

The OHS data set initially included data on over 100,000 individuals, with several hundred variables corresponding to the questions in the questionnaire. For the purposes of this study the sample has been narrowed in a number of ways, and specific variables (as identified above) created from the initial data series. Firstly, the sample was restricted to those between the ages of 16 and 65, which is the life-span over which most people have the potential to be in the labour market. I then created a net monthly earnings variable, and divided this by four to generate net weekly earnings.[5] Net weekly earnings were finally divided by the number of hours worked in the last week of employment, to obtain a proxy for the hourly wage rate. There were a small number of observations for which earnings were not reported correctly, and these were dropped from the sample.

I also generated variables for the number of children fathered or mothered by each individual and living in the same household, for the number of years of education undertaken by each individual, and for the years of potential labour market experience (proxied by age-education-6).[6] I then developed dummy variables for employment, strict and broad labour market participation, and strict and broad unemployment, the methodology for which is included in *appendix 1*. Again, a small number of observations were dropped where responses to certain questions were 'miscellaneous', making it difficult to be sure of correct classification by these distinctions. As a proxy variable for

[4]As already discussed, the wage variable will be an imputed variable derived from a standard human capital function, which itself will require variables relating to education, experience, and location, as well as the actual hourly wage of those in employment.

[5] The net monthly earnings variable was created from two earnings variables already provided in the OHS data set; employee net monthly earnings and self-employed net monthly earnings.

[6] In addition to these, dummy variables for gender, marriage, and rural location were generated directly from the initial data series.

non-labour income, I generated a variable equal to the total per capita income of the household that was not earned by the individual in question.[7]

Finally I made the decision to drop quite a significant number of non-participants from the sample, who were also classified as full-time students. Conceptually this is a tricky issue, and it is not strictly correct to class these individuals as non-participants when they are presumably undertaking study with the aim of contributing their skills to the labour market at a later stage. Their characteristics may therefore hide some of the features that we are trying to uncover of true labour market non-participants. There is however the possibility that a number of students choose to study because they are discouraged in the labour market, and would rather take a job if offered. These students should be classed as labour market participants (i.e. they desire employment), and therefore remain in the sample despite being students.

4.3 A Descriptive Analysis of African Labour Supply

There are good *a priori* reasons to believe that male and female labour markets operate in different ways, and in particular that they will demonstrate different relationships with characteristics such as marriage, age and children. Indeed, the differences in unemployment figures between males and females have already hinted at a dichotomy between the two markets. For this reason I have split the sample into African males and African females, and I will analyse each of these groups separately.

4.3.1 African Males

There are 16402 African males in the final sample, which represents a population of just over 5.7 million when adjusted upwards using the OHS weights. Of this population, 85% are classed as participants in the sense that they are working or express a desire to work. Of these broad participants 70.5% find themselves in employment, while 13.5% are actively seeking work, and a further 16% can therefore be classed as discouraged

[7] For this I made use of the IES data on household income, and the variable was specified as follows: $wnwage_i$ = (weekly household income – weekly earnings of individual i) / household size.

workers.[8] In table 4.1 I have presented summary statistics for the key variables that are thought to influence participation and/or success in finding suitable work. These are the variables that I will use in my econometric analysis, and so it is interesting at this point to make a preliminary examination.

Table 4.1 Mean and Standard Deviations of Key Variables (Males)[a]

Variable	*All Males*		*Non-Working Males*		*Working Males*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Participation[b]	0.85	0.004	0.62	0.009	-	-
Strict Participation[b]	0.71	0.007	0.28	0.010	-	-
Employment[b]	0.60	0.008	-	-	-	-
Weekly Earnings	240.7	6.788	-	-	402.34	9.856
Weekly Hours	26.51	0.361	-	-	44.32	0.214
Hourly Wage	5.94	0.158	-	-	9.92	0.230
Weekly Non-Wage Income	74.61	2.840	74.37	2.136	74.77	4.293
Age	36.40	0.114	34.44	0.189	37.73	0.148
Experience	22.82	0.139	21.40	0.220	23.77	0.185
Education	7.59	0.071	7.03	0.080	7.96	0.090
Rural Location[b]	0.47	0.014	0.53	0.017	0.43	0.015
Married[b]	0.36	0.007	0.21	0.006	0.47	0.011
Children Under 6	0.24	0.007	0.12	0.006	0.33	0.010
Children under 16	0.80	0.017	0.44	0.016	1.04	0.026
Sample Size	16402	-	6866	-	9536	-

[a]Standard Deviations corrected for sample design. Source: OHS 1995
[b]Dummy Variables (1=Yes; 2=No).

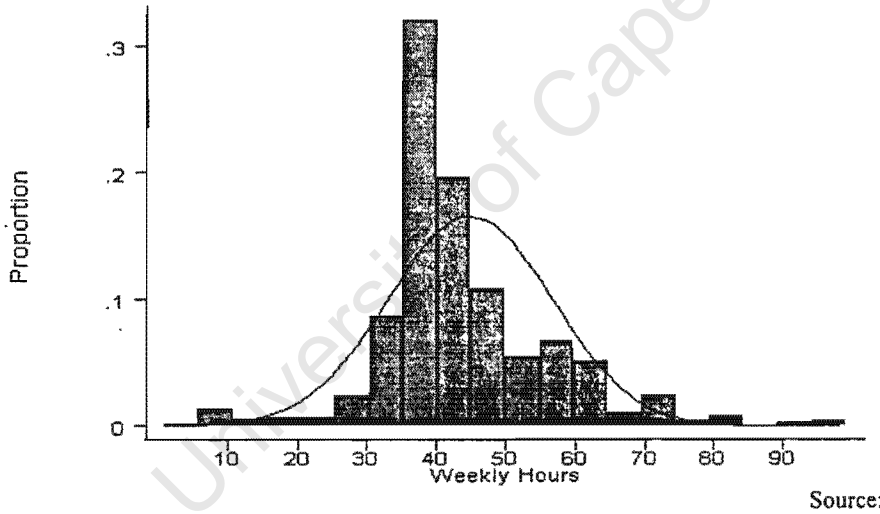
The statistics are fairly self-explanatory, although a number of interesting observations can be made. First of all it is noticeable that the average non-wage income is constant across workers and non-workers, suggesting that this will not be an important determinant of employment. Theoretically non-wage income should influence participation through its effect on the reservation wage, and so this is not a surprising finding given the important distinction between participation and employment in South Africa.[9] The statistics also show that the average age and experience of working males is higher than that of non-working males, as are mean years in education. However, the difference in average education between those with and without jobs (less than a year) is perhaps not as marked as we would expect. Working males are also less likely to live in rural areas and more

[8] Although note again the point made by the ILO (1996:125) and mentioned in chapter two, with regard to interpreting these individuals as discouraged workers.
[9] Non-wage income does have a much higher (more than double) standard deviation among those who are working, however. A possible interpretation of this is to think of two groups of individuals who find employment. At one extreme are those with very low reservation wages (associated with limited alternative sources of income in the household), and at the other are those for whom reasonable household (non-wage) income has afforded them the education to advance more successfully in the labour market.

than twice as likely to be married. This is consistent with the coping mechanism hypothesis discussed in chapter two, whereby the unemployed show a propensity to delay marriage and independent household formation [Klasen & Woolard (1998:17), Wittenburg (1999:6)]. The statistics on the average number of children also reflect such a phenomenon.

The participation, hours and earnings variables are of particular interest, as these will form dependant variables in my econometric analysis. Participation is a dichotomous variable, and we can see that the majority of the population are classed as participants. In contrast the hours and earnings variables are continuous, although they are only observed when an individual is actually employed.

Figure 4.1 The Distribution of Weekly Hours (Males)

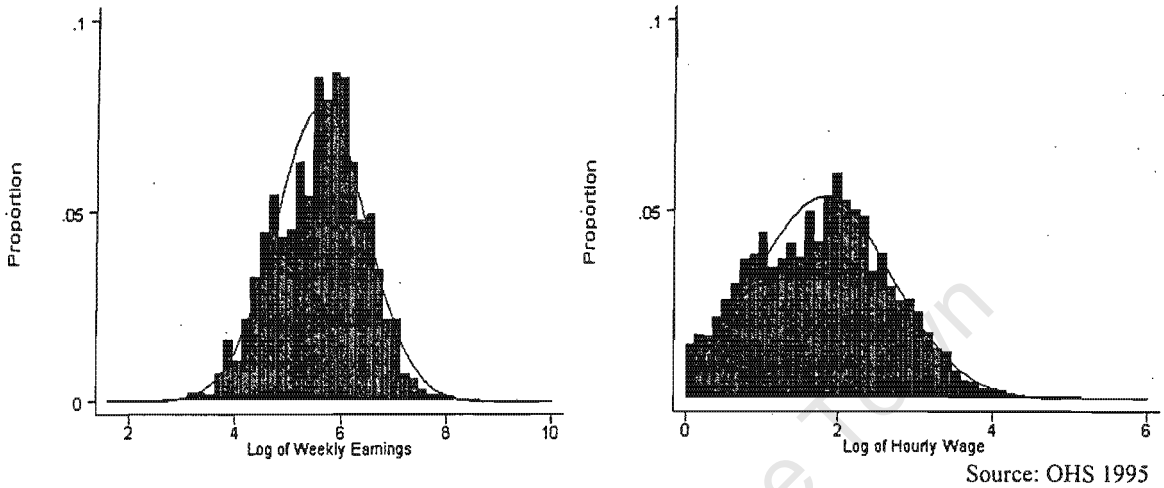


Source: OHS 1995

In figure 4.1 I have drawn the distribution of the weekly hours variable for employed individuals, presented in the form of a histogram. We can see a distinct clustering here just below 40 hours, with other noticeable concentrations on either side. This is perhaps indicative of institutional constraints and conventions on the hours decision, and brings us back to the discussion at the end of chapter three on the assumption that desired hours are attainable. Equally, however, we cannot assume that the workers in these clusters are definitely rationed, as these outcomes may well reflect desired choices. Indeed, the hours variable very broadly follows the shape of a normal distribution, and there does seem to be sufficient variation to allow for an assumption of continuous hours. Another feature of the

distribution is its skewness to the right, with very few observations found below 30 hours. This suggests that part time employment is not an important feature of the male labour market.

Figure 4.2 The Distributions of Log Weekly and Log Hourly Earnings (Males)



Source: OHS 1995

In figure 4.2 I have drawn similar histograms for the log weekly and log hourly earnings, and again we can see a good range of observations. The log of weekly earnings is indeed close to the normal distribution, although when divided by hours it becomes particularly fat in the left-hand tail, illustrating a clustering of low hourly wages. Indeed the tail is cut where the distribution hits the y-axis, with a significant grouping at just above one rand per hour (ln(1)=0).

4.3.2 African Females

The female sample consists of 20510 individuals, corresponding to a population of just over 6.2 million. Of this population there are 60.6% who are broadly classified as participants, only 51% of who have successfully found employment. The remaining 49% of participants are split between active work seekers (20%) and discouraged workers (29%). Summary statistics for the dependant variables and the main variables thought to influence participation, employment and hours are presented in table 4.2.

Table 4.2 Mean and Standard Deviations of Key Variables (Females)[a]

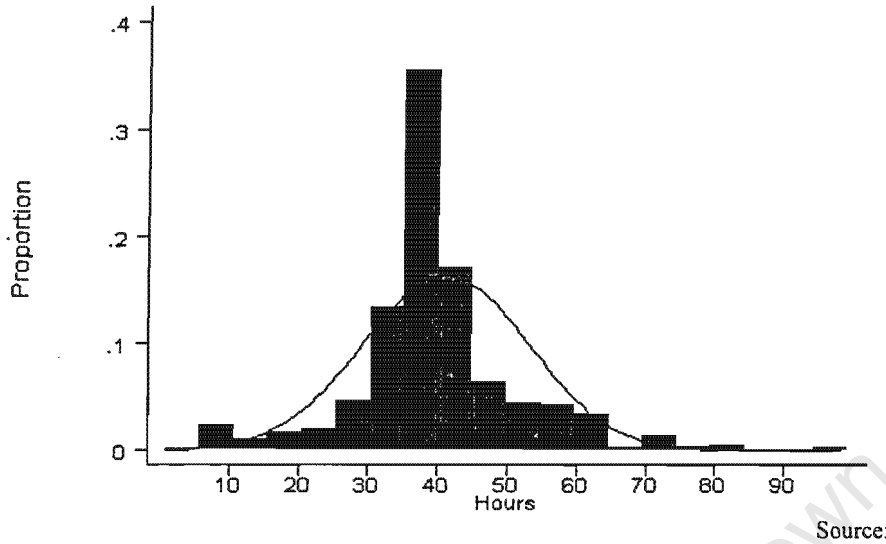
Variable	*All Females*		*Non-Working Females*		*Working Females*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Participation[b]	0.61	0.006	0.43	0.007	-	-
Strict Participation[b]	0.43	0.007	0.17	0.006	-	-
Employment[b]	0.31	0.006	-	-	-	-
Weekly Earnings	100.58	3.409	-	-	323.92	8.009
Weekly Hours	12.72	0.255	-	-	40.96	0.220
Hourly Wage	2.78	0.098	-	-	8.95	0.246
Weekly Non-Wage Income	83.38	2.170	75.90	1.934	99.98	4.435
Age	36.87	0.095	36.59	1.244	37.47	0.147
Experience	23.83	0.121	24.27	1.152	22.86	0.189
Education	7.04	0.068	6.32	0.651	8.62	0.099
Rural Location[b]	0.59	0.013	0.66	0.013	0.43	0.015
Married[b]	0.34	0.005	0.33	0.006	0.35	0.009
Children Under 6	0.40	0.006	0.43	0.008	0.33	0.010
Children under 16	1.25	0.015	1.27	0.018	1.21	0.022
Sample Size	20510	-	143421.	-	6168	-

[a]Standard Deviations corrected for sample design. Source: OHS 1995
[b]Dummy Variables (1=Yes; 2=No).

These figures highlight a number of distinct differences between the male and female labour markets, and would certainly appear to justify a separate analysis of the two. The non-wage income variable is no longer constant in mean across the whole sample, and is significantly higher for workers than for non-workers (again with much higher variation). One possible explanation arises if we link this to education, where we can see a much larger difference in educational achievement between the two groups than for males. On average female workers have more than two years of educational advantage over non-workers, and we can perhaps link this to a greater importance of education, giving those females from a household which has the resources to afford better education an increased chance of finding work.

Like the male sample, female workers are older and more experienced on average than non-workers, and are also far less likely to live in rural areas. The effects of marriage and children are quite different however. The proportions of female workers and non-workers that are married are almost identical, while for males there were greatly increased chances of being married for those that were employed. For females there is also little difference across the sample in the average number of children aged sixteen or below, but working females are less likely on average to have children under the age of six.

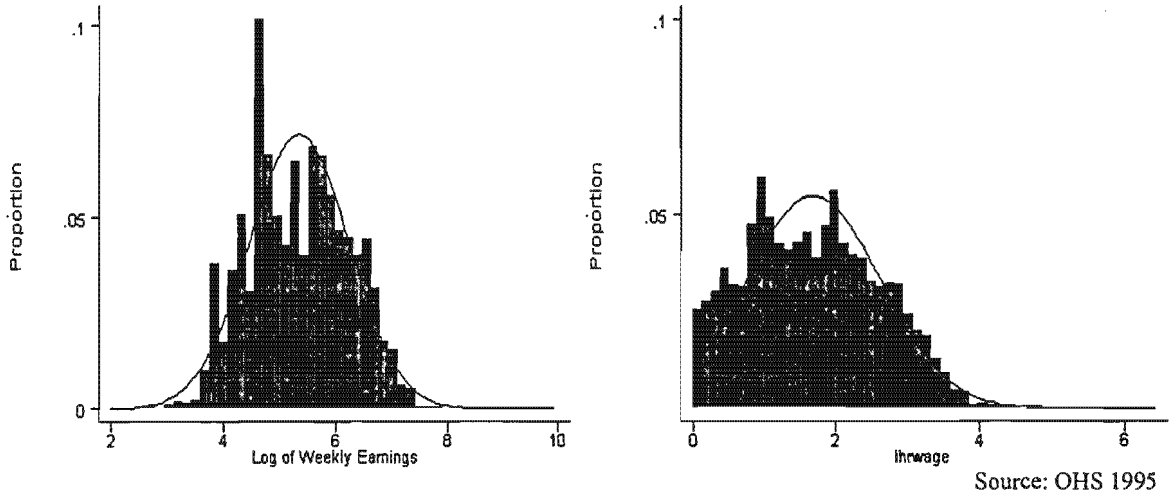
Figure 4.3 The Distribution of Weekly Hours (Females)



Source: OHS 1995

Turning to the dependant variables, we can note that the participation and employment rates are much lower than for the male sample. For those that are employed, the hours distribution is drawn in figure 4.3, where we can see a similar pattern to the male distribution, with a large grouping just under 40 hours. Again there is considerable variation and a 'broadly' normal pattern to the distribution, although the clusters of observations to the left of the 40-hour mark are slightly more significant than for the male sample. Part time employment may be marginally more important in the female labour market therefore.

Figure 4.4 The Distributions of Log Weekly and Log Hourly Earnings (Females)



Source: OHS 1995

Figure 4.4 illustrates the distribution of female log weekly and log hourly earnings, and the pattern appears quite similar to that of male earnings. The distribution of log weekly earnings does not fit the normal distribution as well as for males however, and appears to be wider and flatter. It is also distributed around a lower mean than the male sample, and that female earnings are on average lower than male is clear from the summary statistics. Furthermore, there is a particularly large cluster of observations to the left of the mean, which corresponds to around R100 per week, or R400 per month. This may be the result of some conventional wage for female labour; for example the wage commonly paid to domestic workers. Finally, the distribution is again much fatter in the left-hand tail when weekly earnings are divided through by hours, with a concentration of observations evident at very low hourly wages.

4.4 Awareness of the Potential for Measurement Errors

Before moving on to an econometric analysis using the data and variables summarised in this chapter, it is worth mentioning briefly a few points concerning the potential for measurement error in these variables. Measurement error is a problem that plagues most empirical studies, but it is of particular importance with respect to the variables used in estimating labour supply.

Firstly, data sets often contain direct measures of labour supply only over short periods, such as the week previous to the questionnaire being administered [Killingsworth (1983:98)]. Consequently there is substantial scope for error to creep into the analysis, due to variation in hours over weeks of the year, and inconsistencies between reported hours in the reference week and the more general reporting of earnings. This limitation is particularly relevant for the hourly wage therefore, which in this case is derived from two different (and not necessarily always consistent) questions in the OHS questionnaire.

Secondly, in the common case where wage and labour supply variables are derived from each other, there will be the added problem that measurement errors in the dependant and independent variables may be correlated [Killingsworth (1983:87)]. For example, where the hourly wage is included as an independent variable in the hours function there will be a danger of spurious correlation due to errors of measurement common to the two

variables. This creates a strong argument for using an imputed wage variable, even when estimating a labour supply function over only those who are employed.

Finally, while the most significant scope for error is likely to be in the hours and wage variables, we should also be aware of the limitations of other variables, and especially those that are proxies for a more distinct determinant. In particular, the commonly used experience proxy is very limited in capturing actual labour market experience, and the education variable gives no information as to the quality of education or indeed a more fundamental measure of workplace ability. The correct proxy for non-wage income is also open to debate, and with regard to the measure used in this study there are questions to be asked over the extent to which individuals have access to household income that they do not themselves earn. We should therefore proceed on the premise that the variables have been defined as well as possible, and that while they are likely to form a reasonably good proxy for more accurate measures, we should at least be aware of their potential limitations.

Chapter Five

An Econometric Modelling of African Labour Supply

5.1 Introduction

Previous chapters have taken us through some of the theoretical and policy debate surrounding unemployment, both at a general level and specifically in the economy of South Africa. We have established that unemployment seems to be largely an African phenomenon, and in this chapter I will undertake a detailed econometric analysis of the decisions that Africans make with regard to their labour supply, and the constraints placed on these decisions in the actual market. In estimating facets of the labour supply decision I will broadly follow the techniques discussed in chapter three, while ensuring that the important distinction is made between participation and actual employment.. The emphasis will therefore be on a step-by-step approach to developing an understanding of the factors associated with the labour supply decision.

The first step is to estimate a wage equation, so that imputed wages based on human capital characteristics can be made for those who are not currently employed. Following from this I estimate a model of labour force participation using the probit model for binary dependant variables, and examine the relationship between participation and various characteristics. I then take the sample of participants, and estimate a further probit model to analyse the consequent employment of participants. Finally I examine the decision over how many hours of labour to actually supply. Specifically this involves estimating a joint model of employment and hours (conditional on participation), and then comparing this with a model associated with the traditional neo-classical interpretation of the labour supply decision. Given the likely differences between the two labour markets, each of these steps are carried out separately for males and females.

5.2 Wage Equations

The first step in an econometric analysis of labour supply is to solve the problem of missing wages for those individuals who are either non-participants or unemployed participants. To this end I have experimented with four different regression techniques. The first is a simple OLS estimation of the standard human capital function, based only on those for whom wages are observed. This method is included mainly for purposes of comparison, and is likely to produce biased estimates due to the fact that those observed in employment will not be randomly selected into the sample. Following conventional human capital functions, and to mute the inherent skewness of the raw wage data, my dependent variable is the log transformation of the hourly wage. The variables thought to determine wages are education, experience and location (rural and provincial dummies), with a squared experience term also included to allow for the possibility of a quadratic relationship.[1]

The second model is the simple Tobit model of equation 3.9, estimated by maximum likelihood and using the same variables as the OLS specification. This model essentially corrects for the censoring evident in the data, while assuming that the same group of variables determines both whether the wage is censored and the actual level of the wage. We can therefore interpret the dependent variable as a latent (log) wage, whereby an actual wage is observed only if the latent wage is greater than zero.

The third and fourth models are sample selection models, and correct the estimated wage equation for the fact that individuals with observed wages are not selected at random. The first model uses the Heckman two-step procedure (with corrected standard errors), while the second model estimates the two equations simultaneously by full information maximum likelihood. As discussed in chapter three, this is essentially a generalisation of the Tobit model which is not so restrictive in it's assumptions and allows different variables to influence the selection and wage parts of the model to different degrees.

[1] It is questionable whether these traditionally used variables fully capture the characteristics that are most important in determining earnings. For example, they are limited in transmitting elements of intelligence, motivation, continuity of experience, and quality of education & training [Andersson (1993:63)]. Surveys such as the OHS are not generally conducive to more qualitative measures of individual characteristics, however, and we are therefore restricted to the traditional proxies.

Indeed, it is important for the identification of such models that there is at least one variable that is thought to determine the chances of observation but not the wage outcome.[2] In this case there are three additional variables (children, marriage, and non-wage income) included in the selection part of the model.[3]

These four models were estimated for both males and females, the only difference being a slightly changed specification of the child variable in the participation part of the selection models.[4] The full results of the four regressions (including the selection equation for the last two models) are included in *appendix 3*. In tables 5.1 and 5.2 below I have presented the basic findings of the wage equations.

Table 5.1 Wage Equation Estimates for African Males[a]

Dependant Variable: Log of Hourly Earnings

Explanatory Variable	*OLS Regression[b]*	*Tobit Regression*	*Sample Selection (Two-Step)*	*Sample Selection (FIML)*
Education	**0.110**	**0.162**	**0.090**	**0.097**
	(0.0035)	(0.0039)	(0.0001)	(0.0001)
Experience	**0.029**	**0.158**	**-0.006**	**0.006**
	(0.0030)	(0.0039)	(0.0001)	(0.0002)
Squared	**-0.0002**	**-0.002**	**0.0003**	**0.0001**
Experience	(0.00005)	(0.00007)	(0.000001)	(0.000003)
Rural Dummy	**-0.383**	**-0.183**	**-0.339**	**-0.356**
	(0.0288)	(0.0311)	(0.0008)	(0.0009)
Constant	**0.492**	**-2.288**	**1.419**	**1.093**
	(0.0736)	(0.0881)	(0.0024)	(0.0033)
Lambda	-	-	**-0.577**	**-0.370**
			(0.0006)	(0.0012)
Rho	-	-	-0.85	-0.50
Observations	9536	16402	9536	9536

[a] Significant Coefficients at 95% in bold, at 90% with an *.
[b] Survey Regression corrected for the effects of clustering (all regressions are corrected for weighting).

[2] Theoretically such identifying variables are not necessary, but without them we are relying on the functional form to identify the model [Stata Corporation (1997b:188)].
[3] I have also used age rather than experience in the selection equation, as logically this is perhaps more likely to determine whether or not someone is observed in employment. In practice the variables are highly correlated (due to the derivation of the experience variable) and so the change is a subtle one.
[4] Younger children are likely to have more of an impact on whether females are observed as employed, as childcare still falls largely in the female domain. I have therefore used children below school leaving age (6) as a variable in the female regressions. For males, however, any effect that children may have is more likely to be an incentive effect induced by the need to support them, and thus a broader child definition is

Table 5.2 Wage Equation Estimates for African Females[a]

Dependant Variable: Log of Hourly Earnings

Explanatory Variable	*OLS Regression*[b]	*Tobit Regression*	*Sample Selection (Two-Step)*	*Sample Selection (FIML)*
Education	**0.136**	**0.244**	**0.117**	**0.124**
	(0.0037)	(0.0056)	(0.0002)	(0.0002)
Experience	**0.010**	**0.166**	**-0.006**	0.0002
	(0.0030)	(0.0055)	(0.0002)	(0.0002)
Squared	0.0009	**-0.002**	**0.0003**	**0.0002**
Experience	(0.00007)	(0.00009)	(0.000003)	(0.000004)
Rural Dummy	**-0.167**	**-0.560**	**-0.087**	**-0.120**
	(0.0324)	(0.0435)	(0.0013)	(0.0015)
Constant	**0.278**	**-4.247**	**0.892**	**0.648**
	(0.0909)	(0.1389)	(0.0051)	(0.0074)
Lambda	-	-	**-0.286**	**-0.172**
			(0.0015)	(0.0029)
Rho	-	-	-0.39	-0.23
Observations	6168	20510	6168	6168

[a] Significant Coefficients at 95% in bold, at 90% with an *.

[b] Survey Regression corrected for the effects clustering (all regressions are corrected for weighting).

These equations are of interest in themselves, but their purpose in this study is to develop an imputed wage variable, and so I will not spend a long time analysing the coefficients. It is clear from the results of all four models, however, that for both genders education has a significant and positive influence on the wage rate, and living in rural areas has a strong negative effect. Furthermore, a significant coefficient on lambda in the sample selection models indicates a selection bias in the data, and in particular a negative rho signifies an inverse relationship between the error terms of the selection and wage equations. Predicted wages from estimates that are uncorrected for this bias will therefore be lower than they should be across the population.[5] Indeed, the downward bias in the OLS estimates can be seen in tables 5.3 and 5.4 below, where I have presented some summary statistics for the four different predicted log of wage variables. These have then been transformed into actual imputed wage variables, and these summary statistics are

used (up to age 16).

[5] This is an interesting result, as we would usually expect rho to be positive in a sample selection equation for wages. That it is negative suggests that those who are not observed in employment have unobservable characteristics that are associated with higher wages than those who are observed in employment. This result can be seen as a facet of the unusual unemployment situation in South Africa, and can possibly be linked to the observation in chapter two that unemployment is actually much higher among those with some education than among those with none at all. This perhaps suggests a disproportionately large pool of employment opportunities for those with no education, perhaps associated with the mining sector and/or rural farm-workers.

presented in table 5.5.

Table 5.3 Summary Statistics of Predicted Log Wage (Male)

Variable	*Mean*	*Median*	*Min*	*Max*	*Kurtosis*	*Skewness*
LhatOLS	1.749	1.751	0.104	3.595	2.61	0.01
LhatTOB	0.557	0.560	-1.885	3.207	2.60	-0.009
LhatSTEP	2.113	2.135	0.736	3.670	2.61	-0.11
LhatML	1.983	1.998	0.512	3.641	2.62	-0.07

Observations: 16402

Table 5.4 Summary Statistics of Predicted Log Wage (Female)

Variable	*Mean*	*Median*	*Min*	*Max*	*Kurtosis*	*Skewness*
LhatOLS	1.487	1.510	-0.103	3.322	2.56	-0.03
LhatTOB	-1.070	-1.032	-4.041	2.691	2.56	0.13
LhatSTEP	1.827	1.847	0.457	3.385	2.62	-0.07
LhatML	1.691	1.710	0.234	3.359	2.59	-0.05

Observations: 20510

Table 5.5 Summary Statistic Comparisons for Imputed Wage Variables

Variable	*Male*			*Female*		
	Mean	*Min*	*Max*	*Mean*	*Min*	*Max*
WhatOLS	6.551	1.110	36.423	5.114	0.902	27.708
WhatTOB	2.431	0.152	24.693	0.647	0.018	14.752
WhatSTEP	9.134	2.087	39.236	6.931	1.579	29.509
WhatML	8.086	1.669	38.128	6.130	1.263	28.764
Hrwage	5.415	0	367.013	2.606	0	617.000
Hrwage(>0)[a]	9.314	1.003	367.013	8.665	1.006	617.000

[a] Calculated only for the individuals where wage is greater than zero

It is evident from tables 5.3 to 5.5 that the variables produce quite different estimates of the log wage and wage. The two sample selection equations produce a similar range of estimates and mean, although the maximum likelihood technique (which is theoretically preferred) produces lower wages on average for both males and females. Both however are higher than the simple OLS estimates, as reflected in the negative value of rho.

The Tobit model is perhaps the most interesting, and produces estimates that are considerably different from the other three. Although they show the same broadly normal distribution as the other models[6], the log wage estimates are much lower and this is carried

[6] The kurtosis and skewness statistics are indeed very similar for all models (both male and female), and

through to the imputed wages. This illustrates the fundamental difference between the Tobit model and the others, in that the Tobit is a modelling of the latent wage, whose observation is conditional on it being greater than zero. With such a large number of non-employed individuals we would expect this to be reflected in low average log wage (and therefore wage) estimates. Indeed for females, where around 70% of individuals are observed with zero wages, both the mean and median log wage are unsurprisingly less than zero.

This should not concern us, however, as we are interested in a proxy for the offered wage which is based on the characteristics of individuals and reflects the censored nature of the data. While not a good estimate of an actual predicted wage, the Tobit estimates will provide a relevant proxy for the wage offer of individuals. My choice therefore lies between the FIML sample selection model and the Tobit model. Given my concern with finding a proxy for the imputed wage rather than actual estimates, it is convenient to use the Tobit results with their simple interpretation as estimates that correct for a simple censoring of the data.[7]

5.3 The Participation Decision

In a pure neo-classical modelling of labour supply the next step of the analysis would be to jointly estimate the participation and hours decisions, using either a simple Tobit model, or a more general sample selection equation. However, this can only be correct if there is no distinction to be made between participation and employment, as such a modelling would effectively group the unemployed with true non-participants. I have argued that this distinction is extremely important in South Africa, and that the South African labour market is characterised by rationing in the form of involuntary unemployment. While a neo-classical approach is appropriate if the aim is purely to produce estimates of hours elasticities (corrected for selection into employment), it would miss an important element

with Kurtosis of close to three and very little skewness, they approximate the normal distribution.

[7] This is especially so given that censoring or selection is essentially two-fold in South Africa, first on participation and then on finding employment. It is perhaps better therefore to opt for the simple model and interpretation, given our lack of concern with accurate wage level estimates.

of the labour market along the way.[8] Given the concern of this study with unemployment, a pure neo-classical approach is not appropriate therefore, and the next step in my analysis is slightly different. In this section I exclusively model the decision to participate in the labour market, before going on at the next stage to analyse selection from participation into employment, and finally the hours supplied decision.

I have modelled the participation decision using a probit model, whereby the dependent variable is a dichotomous variable equal to one if the individual is a participant and zero if a non-participant. The model can be interpreted in terms of a latent propensity to participate (P*), which is determined by a number of characteristics and results in observed participation when greater than zero.

Theoretically participation is determined by evaluating the market wage offer against the reservation wage, and the right-hand-side of the model should therefore include the market wage offer relevant to each individual, and factors that are thought to determine the reservation wage. I have thus specified my model to include as independent variables the hourly wage offer, weekly non-wage income, number of children, and dummy variables for age group and marriage. In table 5.6 below I have presented the results of this model, both in the form of probit coefficients and the corresponding marginal probability effects.

The probit coefficients can be interpreted as a change in the underlying propensity to participate for a unit change in the respective dependant variable. Their sign and significance level therefore give an indication of the nature and strength of the relationship between each characteristic and the propensity to participate.[9] Perhaps easier to relate to, however, is the partial change in the probability of participation, as expressed by the marginal probability effects. The marginal effect is the slope of the probability function relating a specific independent variable to the probability of participation, while holding all other variables constant [Long (1997:72)]. However, due to the non-linearity of the cumulative normal probability function, the magnitude of this effect depends on the values

[8] Indeed, at a later stage I will follow the neo-classical procedure in estimating such elasticities.

[9] There is a problem with this interpretation, however, originating from the fact that the variance of P* is unknown. The meaning of a change in the coefficients is therefore unclear, and in particular it will be misleading to compare coefficients across different specifications [Long (1997:70)].

of the independent variables at which the density function is evaluated.[10] A common method, and one that is used here, is to calculate the marginal effect at the mean values of the independent variables, and this needs to be kept in mind when interpreting the coefficients. Finally, it is clear that the small change implied by measuring the slope of the probability function will not be appropriate for dummy variables, which in nature are not continuous. The marginal effects for dummy variables therefore express the change in probability of participation, for a change in the dummy variable from zero to one.

Table 5.6 Probit Estimates of the Participation Decision[a] [b]

Dependant Variable: Labour Market Participation (Broad Definition)

Variable	*Probit Coefficients*		*Marginal Probability Effects[c]*	
	Male	*Female*	*Male*	*Female*
Imputed Wage	**0.126**	**0.451**	**0.026**	**0.172**
(TOBIT)	(0.0170)	(0.0380)	(0.0032)	(0.0142)
Non Wage Income	**-0.0004**	**-0.0003**	**-0.00007**	**-0.0001**
	(0.0001)	(0.0001)	(0.00002)	(0.00004)
Marriage Dummy	**0.295**	**-0.310**	**0.058**	**-0.119**
	(0.0435)	(0.0276)	(0.008)	(0.0106)
Children[d]	**0.074**	**-0.113**	**0.015**	**-0.043**
	(0.0141)	(0.0161)	(0.0029)	(0.0061)
Aged 25-34	**0.307**	**0.194**	**0.059**	**0.073**
	(0.0452)	(0.0357)	(0.0086)	(0.0123)
Aged 35-44	0.098*	-0.069*	0.020*	-0.027*
	(0.0636)	(0.0415)	(0.0125)	(0.0159)
Aged 45-54	**-0.287**	**-0.388**	**-0.066**	**-0.152**
	(0.0639)	(0.0409)	(0.0158)	(0.0161)
Aged 55-64	**-1.075**	**-1.174**	**-0.324**	**-0.438**
	(0.0616)	(0.0468)	(0.0227)	(0.0146)
Constant	**0.681**	**0.301**	-	-
	(0.0410)	(0.0321)		
Observed Prob(Part)	-	-	0.848	0.606
Predicted Prob(Part)	-	-	0.876	0.621
Pseudo R-Squared	0.12	0.13	0.12	0.13
Observations	16402	20510	16402	20510

[a] Significant coefficients at 95% in bold, at 90% with an *.
[b] Robust standard errors, corrected for clustering in sample design.
[c] Calculated at the mean of other independent variables.
[d] Children under 6 for females, and under 16 for males.

There are a number of findings that emerge from the results in table 5.6. Firstly we can see that almost all coefficients are significant, the exception being the second age dummy variable. As would be expected, the wage variable has a positive effect on the probability

[10] See Long (1997:71-73) for a more detailed explanation and mathematical exposition of this.

of participation, although this effect is considerably smaller for males with mean characteristics than for females. Increased wages would appear much more likely therefore to draw females into the labour market than males, a finding which seems logical given the traditional roles of the two genders.

The non-wage income variable is also significant, but with negative sign. Given the suggestion by Klasen and Woolard that the unemployed attach themselves to earning households as a coping mechanism, this negative coefficient might imply that the propensity of these attached unemployed to participate may fall as the household income rises. Such an argument is indeed common in Western economies with respect to welfare benefits and their impact on motivation. The marginal probability effect of the non-wage income variable is tiny, however, for both males and females, and consequently we should not be too concerned by such an argument.[11] This is indeed an unsurprising finding given the generally low levels of household income among Africans.

The dummy variables for marriage and children are significant in the models, and both of these exert a positive (although reasonably small) influence on the probability of participation for males. This again is a logical result, as the incentives for participation are likely to be much larger for males with a spouse and children to support. For females the effect is the opposite, with both marriage and children associated with a lower probability of participation. Furthermore, the magnitude of these marginal effects are greater (in the opposite direction) for the average female than they are for the average male.

The age dummies also produce some interesting results, and indicate a pattern in participation over age that is similar for both males and females. The reference group here is 16-24 year-olds, and we can see that the next age grouping (25-34) is associated with a significant increase in the probability of participation. Subsequent age groups then show a decline in probability, with 35-44 year olds not significantly different from 16-24 year-olds, and the following two age groups exhibiting progressively negative and significant marginal effects. There would thus appear to be an inverted-U-shape relationship between participation of the average individual and age, with a peak in the 24-25 age group.

[11] For example, a huge increase in weekly household income of R1000 is associated in this model with only

Finally, it is worth noting that the model predicts a probability of 0.88 that the average male will participate, and 0.61 for the average female. Both of these are slightly greater than the observed proportions or participants.

5.4 The Ability to Find Employment

Having estimated a model of labour market participation, the next step in my analysis is to try and explain the subsequent selection into employment of those that participate. I have therefore estimated a second probit model of employment for those who are classed as labour market participants. There is of course a possibility of selection bias, stemming from the non-random selection of individuals as labour market participants, and I have therefore incorporated a selection term to correct for this. The selection term is the inverse Mills ratio, which was mentioned in chapter three, and is the ratio of the probability density function to the cumulative distribution function as calculated from the participation probit. This is essentially like performing a Heckman two-step procedure therefore, but where the second model has a dichotomous dependent variable.[12]

The relevant independent variables for this model are those which are thought to determine the chances of being employed, given a desire for employment. Education and experience will therefore be the key variables, and there are also good reasons to include dummy variables for location, especially given the perceived uneven distribution of employment opportunities in South Africa. Employment was therefore regressed on these variables, and the results are presented in table 5.7.

Taking the education and experience variables first, there are some interesting differences that emerge between males and females. As expected education has a positive and significant relationship with employment for both, but for males the effect is considerably lower. For a male with mean characteristics, the results suggest that an extra year of

a 0.07 (or 7%) reduction in the probability of participation for the average male.

[12] We should be aware that the two-step procedure generates inefficient estimators when applied conventionally, and while STATA can correct for this in the standard two-step model it has no corresponding command for the situation where we have two dichotomous dependant variables. However, the errors reported are Huber-White measures of robust standard errors, which are correct in measuring the standard error of $\hat{b}$ (even if there is correlation between the sample selection term and the error) if we were

education is associated with a 0.01 (or 1%) increase in the probability of employment, which is not an enormous amount.* For females, the effect is roughly four times as much.

This suggests that education is considerably more important in influencing the employment opportunities of women, and is maybe a reflection that higher female unemployment has created an extra premium on education. The experience term lends further support to this suggestion, and while this is insignificant for males it shows a significant and positive relationship for females, with a 0.02 increase in probability for the average female.[13] With such high unemployment generally, however, we would expect a similarly high premium on male education, and the magnitude of the difference is perhaps suggestive of other factors at work. One possible suggestion is that the type of work available for females generally requires a higher level of education, and that males therefore have less scope to compete in the labour market on grounds of education. Additionally it is possibly indicative of a generally poor standard of education (hindering the ability of males to distinguish themselves in the labour market), which females are perhaps able to make better use of.

From the location variables it would appear that certain provincial differences have a particularly strong effect on the chances of finding employment. The reference province is the Western Cape, which has the lowest general unemployment rate (see table 2.4). For both males and females, the Eastern Cape and Northern Province are associated with significantly lower probabilities of finding employment. Indeed, for the average male participant the probability of finding employment falls by over 0.2 if he is resident in either of these provinces (compared to the Western Cape). For females the effect in the Eastern Cape is much smaller, but in the Northern Province it is also very high (0.11).

The Eastern Cape and Northern Province were identified in chapter two as areas where unemployment is of particular cause for concern, and this concern seems to be supported by the econometric analysis. As mentioned earlier, these two provinces share a common

---

to repeat the data collection followed by estimation again and again [Stata Corporation (1997b:236-237)].

[13] The squared experience term has a negative (although insignificant) coefficient for the female sample, suggestive of an element of diminishing returns to experience. For males the squared term is significant and positive, but very small. Given the insignificance of the non-squared term however, it is difficult to make any useful interpretation of this.

feature in containing much of the former homeland areas. It is difficult to ignore the link between the two, and it would appear as if this legacy of apartheid has left significant disparities across regions of the new South Africa. These are not the only regions that stimulate concern, however, and the model also suggests that Kwa-Zulu Natal, North West and Mpumalanga are associated with much lower chances for males of finding employment.

Table 5.7 Probit Estimates of the Ability of Participants to Find Employment[a] [b]

Dependant Variable: Employment

Variable	*Probit Coefficients*		*Marginal Probability Effects[c]*	
	Male	*Female*	*Male*	*Female*
Education	**0.034**	**0.095**	**0.011**	**0.038**
	(0.0056)	(0.0060)	(0.0018)	(0.0024)
Experience	0.003	**0.053**	0.001	**0.021**
	(0.008)	(0.0052)	(0.0027)	(0.0021)
Squared Experience	**0.0007**	-0.00007	**0.0002**	-0.00003
	(0.0002)	(0.0001)	(0.00005)	(0.00004)
Rural Dummy	**0.164**	-0.020	**0.053**	-0.008
	(0.0455)	(0.0437)	(0.0146)	(0.017)
Eastern Cape	**-0.652**	**-0.292**	**-0.237**	**-0.008**
Dummy	(0.1169)	(0.1171)	(0.0451)	(0.0174)
Northern Cape	-0.140	-0.033	-0.047	0.013
Dummy	(0.2105)	(0.1665)	(0.0741)	(0.0664)
Free State	-0.098	-0.101	-0.033	-0.040
Dummy	(0.1195)	(0.1176)	(0.0408)	(0.0469)
Kwa-Zulu Natal	**-0.364**	-0.046	**-0.126**	-0.018
Dummy	(0.1155)	(0.1154)	(0.0419)	(0.0460)
North West	**-0.266**	-0.113	**-0.092**	-0.045
Dummy	(0.1210)	(0.1200)	(0.0438)	(0.0478)
Gauteng	-0.200	-0.055	-0.067	-0.022
Dummy	(0.1247)	(0.1209)	(0.0431)	(0.0482)
Mpumalanga	**-0.319**	-0.117	**-0.112**	-0.047
Dummy	(0.1204)	(0.1120)	(0.0445)	(0.0478)
Northern Province	**-0.562**	**-0.277**	**-0.204**	**-0.110**
Dummy	(0.1319)	(0.1264)	(0.0513)	(0.0496)
Selection Term	**1.244**	**0.561**	**0.405**	**0.223**
	(0.0882)	(0.0772)	(0.0294)	(0.0307)
Constant	**-2.058**	**-2.281**	-	-
	(0.1468)	(0.1376)		
Observed Prob(Emp)	-	-	0.706	0.512
Predicted Prob(Emp)	-	-	0.738	0.521
Pseudo R-Squared	0.14	0.12	0.14	0.12
Observations	13690	12158	13690	12158

[a] Significant coefficients at 95% in bold, at 90% with an *.
[b] Robust standard errors, corrected for clustering in sample design.
[c] Calculated at the mean of other independent variables.

The effect of the rural dummy variable is also interesting, as for males it is positive and significant. The effect is also relatively large, and suggests that the chances of finding employment are significantly improved for males living in rural areas (when controlling for other factors). This is slightly surprising, although not altogether without explanation. While table 2.3 suggests that male unemployment is higher in rural areas than urban, there is not nearly as large a difference as for female unemployment, which indeed appears to form much of the urban-rural dichotomy. Furthermore, the five provinces just discussed are predominantly rural provinces, and it is likely that much of the rural effect is captured in these results. It would appear therefore that rural unemployment for males is not such a large problem *per se*, but a problem specific to certain predominantly rural provinces.

Finally, we can also see that the selection term is significant in the model and returns a positive coefficient. This signifies a positive relationship between the error terms of the participation and employment equations, and suggests that the non-participants not included in the employment sample have unobservable characteristics associated with lower chances of employment than participants do. This makes sense, and were we not to correct for selection bias we would therefore obtain predicted probabilities of employment that were biased upwards for the population as a whole.

5.5 Estimating Hours Models of Labour Supply

Having made the distinction between participation and employment and examined the factors associated with these two stages of the labour supply process, I will now move on to investigate the final decision of those that find employment, over how many labour hours to supply. There are a number of ways in which this stage can be approached, and I shall estimate three different models for purposes of comparison.

The first model is a continuation of the step-by-step approach that I have been following, and extends the second probit equation for employment into a joint estimation of employment and hours by Tobit model. Again, there is a selection term included to correct for selection at the participation stage. We can therefore interpret this model as a labour supply equation for participants, which reflects the possible selection bias inherent

in only including participants in the sample, and then corrects hours estimates for censoring in the subsequent ability to find employment.[14] The inclusion of a selection term in a Tobit model is not something that is conventionally done, however, and in particular I am unsure about the correctness of the standard errors and the subsequent efficiency of the estimators.[15]

Consequently, I have estimated a standard neo-classical interpretation of the labour supply model as my second equation, grouping non-participants and the unemployed together. This model uses the entire sample to estimate a Tobit labour supply equation that is corrected for a simple censoring of the hours variable, and is perhaps more likely to give us efficient estimates of hours elasticities.

The third model is then a simple (uncorrected) OLS estimation of the labour supply equation only for those who are observed as employed. It is included to compare the other models with the relationship that is observed among those who are employed, and cannot be used to draw more general inferences about the population parameters because it is likely to suffer from sample selection bias. Tables 5.8 and 5.9 contain the estimated equations for the male and female samples, where the variables thought to influence the hours decision are the wage rate, non wage income, age, number of children, and dummy variables for marriage and rural area.[16]

The first equation shows only the four age dummies as being significant determinants of male hours at the 95% level, with the rural dummy and selection term also showing significance at 90%. For African males therefore, age appears to have a strong association with desired labour hours, with a positive increment as we move through the age groups. It is also noticeable that, although insignificant, the wage term has a negative coefficient and the non-wage income term is positive. This is perhaps contrary to what we would expect.

[14] In a sense, therefore, this is an attempt to achieve the theoretical distinction between participation and employment that is made in the simultaneously estimated 'double hurdle' model, mentioned in chapter 3.
[15] Furthermore, STATA does not allow us to calculate robust standard errors with a Tobit model, and we must therefore be aware of this concern when considering the results.
[16] While there is no solid reason to believe that location will influence the hours supplied decision, I included the rural dummy in case there are different perceptions and conventions with regard to hours

Age also appears to be important for females, although this time there are several other variables that show significance. The wage term is significant and positive, and indeed the wage appears to have a large effect on the Tobit index of desired hours. The non-wage income term is also significant and positive, although much smaller, and the children (under 6) variable shows an expected strong negative effect on desired labour hours. Living in a rural area also appears to be associated with lower hours for females. Again, we can see distinct differences in male and female labour supply responses from these equations, and in particular, the significance of the wage variable for females and not for males suggests that female labour supply in general is more responsive to fluctuations in wages.

Table 5.8 Estimates of Labour Supply Equations (Males) [a]

Dependant Variable: Hours

Variable	*Tobit with Selection (Participants)*	*Neo-classical Tobit (whole sample)*	*Simple OLS (Employed Participants)* [b]
Imputed Wage	-1.679	**2.158**	**-0.208**
(TOBIT)	(1.5496)	(0.1382)	(0.0732)
Non Wage Income	0.007	**-0.005**	-0.00004
	(0.0042)	(0.0019)	(0.0017)
Aged 25-34	**8.434**	**18.100**	-0.1760
	(3.6512)	(0.9528)	(0.5640)
Aged 35-44	**15.236**	**18.302**	0.190
	(1.4570)	(1.0951)	(0.6513)
Aged 45-54	**23.060**	**14.084**	-0.192
	(3.7323)	(1.2040)	(0.6852)
Aged 55-64	**40.665**	0.070	-0.612
	(12.1440)	(1.3181)	(0.7366)
Children Under 16	-0.379	**1.993**	0.092
	(0.9335)	(0.2566)	(0.1171)
Marriage Dummy	3.892	**12.765**	0.166
	(3.630)	(0.7493)	(0.3869)
Rural Dummy	1.103*	-0.582	**1.846**
	(0.5652)	(0.6312)	(0.4356)
Selection Term	26.445*	-	-
	(14.629)		
Constant	-29.869	**-7.751**	**44.139**
	(18.542)	(0.9292)	(0.6309)
Observations	13690	16402	9536

[a] Significant coefficients at 95% in bold, at 90% with an *.
[b] Survey regression corrected for the effects of clustering.

worked in rural and urban areas.

The neo-classical Tobit specification models the participation and hours decisions jointly over the whole sample, correcting for censoring of the hours variable due to non-participation. The interpretation of the dependant variable is therefore of a latent measure of desired hours, with desired hours only becoming observed when they are greater than zero. Where is falls short is that it does not account for the possibility of unemployment, and assumes that censoring will only occur through variables associated with participation. It is nevertheless interesting to examine the results that this neo-classical assumption produces, and it is indeed likely that a simple correction for censoring will produce fairly reliable estimates of overall supply elasticities.

Table 5.9 Estimates of Labour Supply Equations (Females) [a]

Dependant Variable: Hours

Variable	*Tobit with Selection (Participants)*	*Neo-classical Tobit (whole sample)*	*Simple OLS (Employed Participants)* [b]
Imputed Wage	**8.864**	**11.614**	-0.187
(TOBIT)	(1.2634)	(0.4233)	(0.1155)
Non Wage Income	**0.005**	0.0006	**0.003**
	(0.0021)	(0.0023)	(0.0012)
Aged 25-34	**19.662**	**23.432**	0.477
	(1.2811)	(1.4311)	(0.6148)
Aged 35-44	**26.023**	**27.078**	**1.348**
	(1.3384)	(1.5261)	(0.6511)
Aged 45-54	**27.120**	**21.564**	0.411
	(1.7553)	(1.6454)	(0.6953)
Aged 55-64	**31.554**	-1.123	1.305
	(3.0796)	(1.9143)	(0.9078)
Children Under 6	**-4.734**	**-5.712**	-0.243
	(0.6783)	(0.6677)	(0.2796)
Marriage Dummy	0.184	**-5.750**	**-1.492**
	(1.0644)	(0.9142)	(0.3599)
Rural Dummy	**-2.411**	**-11.982**	0.501
	(0.8149)	(0.9282)	(0.4657)
Selection Term	-7.527	-	-
	(3.1632)		
Constant	**-9.496**	**-36.802**	**40.466**
	(3.3408)	(1.4787)	(0.6615)
Observations	12158	20510	6168

[a] Significant coefficients at 95% in bold, at 90% with an *.
[b] Survey regression corrected for the effects of clustering.

The results do indeed appear broadly consistant with those from the first equation,

although more coefficients are now significant. For both males and females the coefficient on wages is positive and significant, with the effect again much larger for females. When weighted for the mean values of the wage rate and hours, these translate into elasticities (on the desired hours index) of 0.12 for males and 0.18 for females. The effect of other household income on hours is negative and very small for males, while for females it is insignificant. The child and marriage variables illustrate once again one of the fundamental differences between male and female labour supply, and are significantly positive for males while significantly negative for females. The rural dummy variable is also significant in the female equation, and reaffirms the suggestion (from the first equation) that rural females are associated with lower desired hours than urban inhabitants. Finally, the age dummies indicate in both cases that the middle three age groups are associated with significantly higher desired hours than the youngest and oldest groups.[17]

Finally, a comparison with the OLS estimates for the sample of workers is interesting. The OLS estimates can only be interpreted with respect to the sample who are actually working, and as such are not terribly useful for deriving labour supply elasticities and effects across the population. However, the results do tell us something about the nature of the relationship between hours and various factors for workers. The most striking feature of both equations is the insignificance of most of the variables, indicating that most of the traditional variables are not associated with hours for the working sample. One possible interpretation of this is that it is a reflection of the institutional constraints inherent in the labour market, and that once individuals are in employment, they have very little scope to vary their hours. Indeed, in a high unemployment economy characterised by excess supply of labour, such an explanation is especially feasible.

There are a couple of exceptions to this in the equations, noticeably the wage rate (which has a negative relationship) and rural dummy variable for males, and the marriage and second age dummy for females. This perhaps indicates that male workers in rural areas are expected to work for longer hours than urban workers, and that low-wage male

[17] We should note when interpreting these coefficients that they refer to the combined effect of participation and actual hours worked on the latent desired hours variable. It is useful and indeed possible to decompose this effect using a method developed by McDonald and Moffitt [Long (1997:210)], although unfortunately this is not possible in the statistical programme that was available.

workers in general work longer hours than higher-wage workers. Married female workers, meanwhile, are likely to work fewer hours that those who are not married. I will now carry these results into the final concluding section, where I will draw together the various chapters and summarise some of the conclusions that can be drawn from the overall study.

Summary and Conclusions

Unemployment is perhaps the key economic and social issue in South Africa at the end of the 1990s. Indeed, finding tenable solutions to the problem, which will allow the simultaneous achievement of the dual goals of poverty and inequality reduction, will be crucial for the success of the economy as we move into the new century. It is evident from the figures presented in chapter two that present unemployment rates in South Africa are worryingly and unacceptably high, at around 29% in a broad sense and 16% when only including active work-seekers. Furthermore, they are distributed disproportionately among different population groups within the country. The big question, and the focus of significant debate among economists, politicians and the general population, is what can be done to ease this situation while at the same time meeting other important economic and social goals.

The unemployment debate to date has been varied, encompassing many aspects of the incidence and nature of unemployment in South Africa. In particular, and in line with the more general international debate on unemployment, it's main focus has been on the labour market, and specifically on issues of market flexibility. There has been general agreement for some time that certain reforms to the labour market are required, but much uncertainty as to the type of reforms which are necessary to avoid potential adverse effects on inequality and poverty. Consequently, there have been various studies investigating certain facets of the labour market, in order to clarify the situation and provide for more informed policy decisions. In this study I have undertaken an investigation of South African unemployment from a labour supply perspective, with the aim of contributing to this debate through providing greater insight into the workings of the labour market on the supply side.

Following an introductory first chapter, the second chapter of the study presented a detailed breakdown of South African unemployment figures, followed by a discussion of certain elements of the unemployment debate which arise from these figures. From this analysis we have seen that unemployment is by no means distributed evenly among the

population. In particular, unemployment in South Africa is a greater problem for females than for males, and a much greater problem for Africans than for any other population group. Given that Africans make up around 70% of South Africa's potential labour force, it is clear that solutions to the African unemployment problem are vital if progress is to be made.

It is also apparant from the figures in chapter two that unemployment varies significantly by location, with the rural market appearing worse at providing jobs than the urban market, a factor also reflected in generally higher unemployment in rural provinces. As Klasen and Woolard suggest, however, a proportion of this rural unemployment may in fact be disguised urban unemployment, and a consequence of the coping strategies of the unemployed. This breakdown also highlighed two provinces, the Eastern Cape and the Northern province, where unemployment is an exceptionally large problem, with the broad rate in excess of 40%. These provinces are alike in that they are predominantly rural, and contain significant areas of the former African homelands. The scale of the problem persisting in these regions indicates the need for a special policy focus on labour market problems and opportunities in these provinces.

A further aspect of unemployment to arise from chapter two is its profile across age groups. The figures suggest that unemployment is very much a youth phenomenon, with broad unemployment rising to a staggering 65% (45% narrow) among Africans between the ages of 16 and 24. Again, this facet forms an important element of the debate and is especially linked to arguments concerning the inability of wages to fall among young workers, and the more general inflexibility of the labour market.

From chapter two we obtained a broad picture of the unemployment problem in South Africa and an insight into some of the key issues that underpin the policy debate. What is most clear is that unemployment is primarily (although certainly not exclusively) a problem of the African population group, and of the young among this group. This point was carried into the next chapters, where the analysis of labour suppy was focussed on the African population group.

Chapter three critically discussed the theory and econometrics behind the modelling of labour supply, and identified a number of conventionally used econometric techniques for overcoming the problems that such a modelling entails. Chapter four then provided an introduction to the October Household Survey dataset, and presented summary statistics for characteristics of the African population that are of interest with respect to the modelling of labour supply. This descriptive analysis was conducted seperately for males and females, as it was decided that there were good *a priori* reasons to believe that the two labour markets operate in very different ways.

The analysis of chapter four was largely a prelude to the econometrics of the next chapter, but it uncovered some important features of the data with respect to differences in characteristics between workers and non-workers, and across genders. In particular, the difference in average years of education between male workers and non-workers is not as marked as we might expect, and much less than that for females. Furthermore, while the non-wage income variable is fairly constant in mean across the male sample, it is considerably higher for females in the worker category. These two observations together perhaps suggest that the importance of education for females gives a labour market advantage to those from families who can afford access to better education, an advantage that is not evident for males.

The statistics also corresponded with some of the observed differences in unemployment rates from chapter two, and showed links with several of the theoretical ideas unearthed in the debate around these figures. For example, they showed that working males are less likely to live in rural areas than non-working males, and more than twice as likely to be married. These are consistant with the rural/urban unemployment figures, and the coping mechanism hypothesis of Klasen & Woolard respectively.

Chapter five provided the bulk of the analysis, in the form of a step-by-step econometric modelling of the labour supply decision, with respect to the dependent and independent variables presented and discussed in chapter four. The first step of this analysis was to estimate a human capital function so that an imputed wage could be given to each

individual in the sample. Four alternative methods of estimation were used, and an imputed wage variable was then constructed from the Tobit specification.

Next the participation decision was analysed, using a probit model for dichotomous dependent variables. The imputed wage, non-wage income, marriage, children and age all displayed significant relationships with participation, and two main conclusions can be drawn from this stage. Firstly, wages are an important determinant of participation in the African labour market, and more so for females than for males. Consequently, we can say that increased wages are more likely to draw the average female into the labour market than the average male. At the same time, however, policies which result in falling wages are likely to reduce the participation of females still further relative to males. Given the acknowledged importance to the development process of female participation in the labour market, this is something which should be seriously considered when formulating labour market policy.

Secondly, the relationship between participation and the non-wage income variable is negative, indicating that the probability of participation falls as alternative household income increases. This effect appears to be small however, and suggests that as the African population become more affluent the reduction in labour force participation should not be excessive. It also has implications for the theory espoused by Klasen and Woolard with regard to household formation strategies. They suggest that a major coping strategy of the unemployed is to adapt household structures, possibly to the extent where the unemployed may put an upward pressure on wages through their effect on raising the sufficiency requirements of the household earner. A negative relationship between non-wage income and participation suggests that as households become richer, the unemployed among them may have less desire to participate. Such coping strategies may therefore become more permanent states of affair if wages among workers rise too much. However, we have noted that this effect is small among the African population, and thus this process of events should not be seen as too much of a concern.

The next set of conclusions to be drawn relate to the second phase of the labour supply process, that of the ability to find work once the decision on participation has been made.

The results again derive from a probit model, and there are several important findings. First of all, education naturally appears to be a significant and positive factor in determining the ability of participants to find employment, although the probability effect is almost four times stronger for the average female than for the average male.

Indeed the influence of education is surprisingly low among males. This may be as a result of relative homogeneity in educational background among the African population, and there is a strong possibility that additional labour market skills (not reflected in years of education) play an important role. However we would then expect some of these other skills to be picked-up in the experience variable, and this does not appear to be the case. The indication may therefore be that the selection of participants into employment is more of a random occurance, and perhaps strongly influenced by factors such as location, extensiveness of search, and even luck. This is not a surprising conclusion given that low average educational achievement would tend to confine African males in many cases to manual and relatively unskilled labour, of which there is only so much.

It is especially interesting that high rates of unemployment appear to have resulted in premia on education and experience (in terms of employment) for women, but not for men. Perhaps this tells us something about the type of employment opportunities that are potentially available for women and men, with female opportunities more likely to require a degree of formal education. Alternatively, it may be indicative of the persistance of the traditional role of females within the household, whereby educated females are more likely to persist in actively searching for employment, as well as undertaking a more informed process of search.

For females therefore, increased educational opportunities would appear likely to have a sizeable positive effect on the ability to find employment, a conclusion which supports the policy focus on female education emphasised by much of the development literature. For males, education does not seem as important, although this does not necessarily mean that male education is an unimportant focus for policy. As mentioned above, it is likely that the reason for the lack of importance of education in determining male employment is

precisely because of it's low level, confining many African males to unskilled jobs, and thus relegating education to a factor of little importance.

The ability of males to find employment does appear to be strongly influenced by location however. In particular, the Eastern Cape and Northern Province are areas where the chances of finding employment fall by more than 20% from those in the Western Province. Other rural provinces, including Kwa-Zulu Natal, Mmpumalanga, and North West, are also associated with a marked reduction in the probability of finding employment. Moreover, a positive coefficient on the rural dummy variable suggests that finding unemployment in rural areas is not such a problem *per* se, but a problem of certain rural provinces. This ties in especially well with the household adaption theory of Klasen and Woolard, as there is a strong case to be made that work-seekers return for support to households particularly in the former homeland areas. Ironically, these households are often very poor already, and this strategy is only likely to exacerbate the poverty in these areas.

Surely therefore there is a need to focus policy on the potential for developing stronger labour markets in these areas, that will provide more opportunities for workers who return here for family support, and also for the families already in these areas. An alternative focus could be on schemes to ease the plight of workseekers in the urban areas, where there is arguably greater employment potential. Housing schemes and search advice centres, for example, could encourage more people to remain in areas where opportunities are greater, rather than returning to family support in regions with poorer prospects. A dual approach to both labour markets is perhaps best, and should even out the distribution of unemployment and dispell some of the discouragement prevalent in both areas.

The final stage of the labour supply analysis was to estimate a model of hours supplied. This was undertaken both as a continuation of the step-by-step approach, and as a complete neo-classical Tobit model of the whole process. A simple OLS model was also estimated to compare the relationships with those observed among the employed. The first specification highlighed again the differences between the male and female labour markets, in particular with respect to the wage and non-wage income variables. Both of

these appear to be considerably more important in influencing female labour supply, which is consistant with the idea that traditional family roles are still an important factor. The continuation of the step-by-step approach was slightly dissapointing in it's final stage, however, due in part to a lack of confidence in the standard errors arising from uncertainty around the technique. This suggests the need to think more carefully about the econometric problems that this stage of the modelling imposes, and in particular to consider the simultaneous estimation of a 'double-hurdle' specification as mentioned in chapter three and outlined in appendix 2. These are areas where further work in the South African context would prove beneficial.

In chapter three I underlined the limitations of a neo-classical modelling of labour supply in an economy characterised by rationing. Such a modelling is nevertheless commonly used in the labour supply literature, and it is interesting to consider the results that such a modelling produces in South Africa. I used the Tobit specification to estimate this model, which gives coefficients for the combined effects of participation and the hours decision on the latent desired hours variable. In this context therefore, participation and ability to find employment are in effect bundled together. The resulting desired hours elasticities with respect to the imputed wage prove to be 0.12 and 0.18 for males and females respectively, controlling for factors such as marriage, children and age.

On a general level this indicates that labour supply in South Africa should respond positively to increased wages, although the response will be inelastic. It would appear that, at this stage of development, African labour in general is on the upward sloping part of it's labour supply curve. Furthermore, the combined effect on desired hours of non-wage income is very small (negative) for males, and insignificant for females. These findings together suggest that stable or slightly rising wages in the African labour market should have a small positive effect on labour supply, and not be of concern in dampening the desire to supply labour. Moreover, policies which result in falling wages may impart a negative overall effect on labour supply, especially among females.

Given these findings it would be interesting to extend this work, and a starting point may be a decomposition of the effects of the Tobit model. This would give us a neo-classical

perspective of the relative extent to which wages and other factors influence the participation and hours decisions, and a comparison could then be made with a double hurdle specification which allows rationing to play a role. This study of the supply-side of the African labour market should hopefully provide a useful and relevant starting point for such further analysis, and the conclusions which I have drawn may indeed have important relevance for the design of policy towards the labour market. At least, it provides a more detailed picture of the workings of the supply side of the market, and of the interaction between demand and supply at the employment determination phase.

Bibliography

Andersson, PA (1993). 'Labour Market Structure in a Controlled Economy: The Case of Zambia.' Ekonomiska studier no. 37, Gothenburg: Department of Economics, Gothenburg University.

Archer, S (1992). 'Strategic Choices in the Labour Market', in R.Schrire (ed.) *Wealth or Poverty? Critical Choices for South Africa.* Cape Town: Oxford University Press

Beharrell, A. (1992). *Unemployment and Job Creation.* London: Macmillan Publishing.

Blanchard, O. & Jimeno, (1995). 'Structural Unemployment: Spain versus Portugal.' *American Economic Review*, vol. 85, pp. 212-219.

Blinder, A. (1988). 'The Challenge of High Unemployment.' *American Economic Review*, vol. 78, Papers and Proceedings, pp. 1-15.

Breen, R. (1996). 'Regression Models: Censored, Sample-Selected, or Truncated Data.' Sage University Paper no. 111, London: Sage Publications.

Cowling, K. & Sugden, R. (1994). *Beyond Capitalism*, London: Pinter Publishers.

Clark, A.E. & Oswald A.J. (1994). 'Unhappiness and Unemployment.' *Economic Journal*, vol. 104, pp. 648-659.

Deaton, A. (1997). *The Analysis of Household Surveys, A Microeconometric Approach to Development Policy*, Baltimore: John Hopkins University Press.

Dilnot, A. & Duncan, A. (1992). 'Thinking About Labour Supply.' *Journal of Economic Psychology*, vol. 13, pp. 687-713.

Greene, W. H. (1990). *Econometric Analysis*, New York: Macmillan Publishing.

Hamilton, L. C. (1998). *Statistics with Stata 5*, Pacific Grove: Duxbury Press.

Heckman, J. J. (1993). 'What Has Been Learned About Labour Supply in the Past Twenty Years?' *American Economic Review*, vol. 83, Papers and Proceedings, pp. 116-121.

Hofmeyr, J. F. (1998). 'Segmentation in the South African Labour Market.' University of Natal, Draft Paper.

ILO (1996). *Restucturing the Labour Market: The South African Challenge* (Standing, G., Sender, J. & Weeks, J. eds.). Geneva: International Labour Office.

Jackman, R. (1992). 'An Economy of Unemployment?', in E. McLaughlin (ed) *Understanding Unemployment, New Perspectives on Active Labour Market Policies*. London: Routledge.

Killingsworth, M. R. (1983). *Labour Supply*. Cambridge: Cambridge University Press.

Klasen, S. & Woolard, I. (1998). 'Unemployment, Household Formation, Poverty, and Nutrition in South Africa.' University of Munich and University of Port Elizabeth, Draft Paper.

Klasen, S. & Woolard, I. (1999). 'Levels, Trends and Consistency of Employment and Unemployment Figures in South Africa.' *Development Southern Africa*, vol. 16, pp. 3-35.

Layard, R., Nickell, S. & Jackman, R. (1993). *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford: Oxford University Press.

Lazaro, N., Molto, M. & Sanchez, R. (1997). 'Women's Labour Force Participation and Part Time Work in Spain.' *Labour*, vol. 11, pp. 449-468.

Leibbrandt, M. & Woolard, I. (1999). 'A Comparison of Poverty in South Africa's Nine Provinces.' *Development Southern Africa*, vol. 16, pp. 37-54.

Leibbrandt, M. V., Woolard, C. D. & Woolard, I. D. (1996). 'The Contribution of Income Components to Income Inequality in South Africa: A Decomposable Gini Analysis.' LSMS Working Paper no. 125, Washington: The World Bank.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*, California: Sage Publications

Macroeconomic Research Group (MERG). (1993). *Making Democracy Work: A Framework for Macroeconomic Policy in South Africa.* Cape Town (UWC): The Centre for Development Studies.

Maglad, N. A. (1998). 'Female Labour Supply in Sudan.' African Economic Research Consortium Special Paper no. 30, Nairobi: AERC.

Malinvaud, E. (1984). *Mass Unemployment.* New York: Basil Blackwell.

McLaughlin, E. (1992). 'Towards Active Labour Market Policies: An Overview', in E. McLaughlin (ed) *Understanding Unemployment, New Perspectives on Active Labour Market Policies.* London: Routledge.

Michie, J. (1997). 'Developing the Institutional Framework: Employment and Labour Market Policies', in J. Michie & V. Padayachee (eds.) *The Political Economy of South Africa's Transition.* London: Dryden Press.

Ministry of Finance. (1996). 'Growth, Employment and Redistribution, a Macroeconomic Strategy.' Government White Paper, Pretoria: Ministry of Finance.

National Institute of Economic Policy (NIEP). (1996). 'From the RDP to GEAR: The Gradual Embracing of Neo-Liberalism in Economic Policy.' NIEP, Discussion Paper.

Palachek, S.W. & Siebert, W.S. (1993). *The Economics of Earnings*. Cambridge: Cambridge University Press.

Sahn, D. E. & Alderman, H. (1988). 'The Effects of Human Capital on Wages, and the Determinants of Labor Supply in a Developing Country.' *Journal of Development Economics*, vol. 29, pp. 157-183.

Schultz, T. P. & Mwabu, G. (1998). 'Labor Unions and the Distribution of Wages and Employment in South Africa.' *Industrial and Labor Relations Review*, vol. 51, pp. 680-703.

Stata Corporation. (1997a). *Stata User's Guide; Release 5*, Texas: Stata Press

Stata Corporation. (1997b). *Stata Reference Manual; Release 5*, Texas: Stata Press

Stoneman, C. (1994). 'The World Bank, Income Distribution and Employment: Some Lessons for South Africa', in K. Cole (ed.) *Sustainable Development for a Democratic South Africa*. London: Earthscan Publications.

Winter, C. (1998). 'Women Workers in South Africa: Participation, Pay and Prejudice in the Formal Labor Market.' Washington: The World Bank, Draft Paper.

Wittenburg, M. (1999). 'Job Search and Household Structure in an era of Mass Unemployment: A Semi-Parametric Analysis of the South African Labour Market.' University of Witwatersrand, Draft Paper.

World Bank (1995). *World Development Report 1995*. Oxford: Oxford University Press.

Zabel, J.E. (1993). 'The Relationship between Hours of Work and Labor Force Participation in Four Models of Labor Supply Behaviour.' *Journal of Labor Economics*, vol. 11, pp.387-415.

Appendix One

Classifying Participants

An individual is classified as a labour market participant (in a broad sense) if he/she satisfies the following:

- Will accept a job if offered [Q3.28], and....
- Does not list 'sought or underwent training' as the primary method of seeking work.
- Does not give 'scholar or student' as a reason for not seeking work in the last 7 days [Q3.27] and at same time answer 'yes' to student question in part 2 of the questionnaire [Q2.11].
- Does not give 'Illness, invalid, disabled or unable to work' as reason for not seeking work and at the same time classify themself with a disability in part 2 of the questionnaire [Q2.28].
- Does not give 'housewife' as a reason for not seeking work, and at the same time classify themself as a housewife in the activities question [Q3.1].
- Does not give 'retired' as a reason for not seeking work and at the same time classify themselves as retired in the activities question.

Classifying Employed

An individual is then classified as employed if he/she satisfies the following:

- Categorised as 'working full time', 'working part time', or 'with a job but absent from work' in the activities question [Q3.1].
- Does not give 'off season activity' or 'reduction in economic activity' as a reason for not working in the last week [Q3.27].

- Does not answer question 3.26 on whether he/she would accept a suitable job. The design of the questionnaire means that employed people should not answer this question.

Or if he/she;

- Lists planning and undertaking preparations for starting his/her own business as the primary method of seeking employment in the last four weeks [Q3.30].

Classifying Unemployed

- The broad unemployed are then made up of all broad participants, less those in employment.
- The strict unemployed are the broad unemployed, less those who answer 'nothing, but still wants work' when asked what they have done to find work in the last four weeks [Q3.30].

Appendix Two

The Double Hurdle Model

The double hurdle model is generalised Tobit model, which is specified in a way that allows for unemployed work-seekers to be considered as labour market participants. They are therefore distinguished from those non-participants who are not seeking work. Following Andersson (1993) the model is specified as follows:

$$P_i^* = \mathbf{X}_i'\alpha + u_i$$

Where; $P_i = 0$ if $P_i^* \leq 0$

$P_i = 1$ if $P_i^* > 0$

And;

$$H_i^* = \mathbf{Z}_i'\beta + e_i$$

Where; $H_i = H_i^*$ if $P_i = 1$ and $H_i^* > 0$

H_i not observed if $P_i = 0$ and $H_i^* > 0$, or if $H_i^* \leq 0$

Essentially this is an extension of the generalised Tobit model of equations 3.10 and 3.11, with the initial latent variable P* now interpreted as the propensity to actually participate, rather than the propensity to have a job. This is accomplished by adding additional conditions to the second part of the model, stating that for H* to be observed, not only must P* be greater than zero, but so must H*. The latent variable H* then becomes interpreted as the propensity to find a job once participating in the labour market. The two stages have therefore been separated from each other.

Appendix Three

Full Regression Results from Wage Equations

1. *Male OLS Regression*

```
pweight:  w                                    Number of obs    =      9536
Strata:   <one>                                Number of strata =         1
PSU:      cluster                              Number of PSUs   =      1952
                                               Population size  =   3418998
                                               F( 12,    1940)  =    179.09
                                               Prob > F         =    0.0000
                                               R-squared        =    0.3709
------------------------------------------------------------------------------
 lhrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .1102238   .0035022    31.473   0.000     .1033554    .1170921
   exper |   .0289422   .0030202     9.583   0.000      .023019    .0348653
 expersq |   -.000206   .0000526    -3.920   0.000    -.0003091    -.000103
   RURAL |  -.3825504   .0287689   -13.297   0.000    -.4389714   -.3261294
   PROV2 |   .1106088   .0643525     1.719   0.086    -.0155982    .2368157
   PROV3 |  -.0058127   .1041741    -0.056   0.956    -.2101168    .1984914
   PROV4 |  -.2742853   .0646195    -4.245   0.000    -.4010159   -.1475547
   PROV5 |   .2180253   .0622133     3.504   0.000     .0960139    .3400368
   PROV6 |   .0995374   .0694802     1.433   0.152    -.0367257    .2358006
   PROV7 |   .2223613   .0613654     3.624   0.000     .1020127    .3427099
   PROV8 |   .1490629   .0724448     2.058   0.040     .0069855    .2911402
   PROV9 |   .4870552   .0785106     6.204   0.000     .3330818    .6410286
   _cons |   .4924242   .0736495     6.686   0.000     .3479841    .6368642
------------------------------------------------------------------------------
```

2. *Female OLS Regression*

```
pweight:  w                                    Number of obs    =      6168
Strata:   <one>                                Number of strata =         1
PSU:      cluster                              Number of PSUs   =      1761
                                               Population size  =   1937445
                                               F( 12,    1749)  =    176.25
                                               Prob > F         =    0.0000
                                               R-squared        =    0.3813
------------------------------------------------------------------------------
 lhrwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |   .1355163   .0037418    36.217   0.000     .1281774    .1428552
   exper |   .0099373    .003478     2.857   0.004     .0031157    .0167588
 expersq |   .0000873   .0000656     1.331   0.183    -.0000413     .000216
   RURAL |   -.166975   .0324173    -5.151   0.000    -.2305555   -.1033945
   PROV2 |   .0356457   .0764961     0.466   0.641    -.1143871    .1856784
   PROV3 |  -.0794252   .1079114    -0.736   0.462    -.2910732    .1322228
   PROV4 |  -.3221862   .0768693    -4.191   0.000    -.4729509   -.1714215
   PROV5 |    .062763   .0743413     0.844   0.399    -.0830435    .2085696
   PROV6 |   .0342542    .082418     0.416   0.678    -.1273934    .1959017
   PROV7 |   .1832304   .0722561     2.536   0.011     .0415136    .3249473
   PROV8 |   .0611172   .0832703     0.734   0.463    -.1022019    .2244363
   PROV9 |   .3553925   .0886344     4.010   0.000     .1815527    .5292322
   _cons |   .2776885   .0909209     3.054   0.002     .0993642    .4560127
------------------------------------------------------------------------------
```

3. *Male Tobit Regression*

Number of obs = 16402
chi2(12) =3846.82
Prob > chi2 = 0.0000
Log Likelihood = -23125.58 Pseudo R2 = 0.0768

lhrwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.1623678	.0038512	42.161	0.000	.1548191	.1699165
exper	.158478	.0039387	40.236	0.000	.1507577	.1661984
expersq	-.0022667	.0000705	-32.137	0.000	-.0024049	-.0021284
RURAL	-.1827891	.0311443	-5.869	0.000	-.2438353	-.1217429
PROV2	-.7726769	.0729409	-10.593	0.000	-.915649	-.6297048
PROV3	-.2331482	.1389581	-1.678	0.093	-.5055211	.0392247
PROV4	-.2848527	.0768242	-3.708	0.000	-.4354365	-.1342689
PROV5	-.3064796	.0690609	-4.438	0.000	-.4418464	-.1711128
PROV6	-.1345537	.0728988	-1.846	0.065	-.2774433	.0083359
PROV7	.1349511	.0666352	2.025	0.043	.0043388	.2655634
PROV8	-.0762833	.0766335	-0.995	0.320	-.2264932	.0739267
PROV9	-.2752886	.0776849	-3.544	0.000	-.4275595	-.1230177
_cons	-2.287686	.0881274	-25.959	0.000	-2.460425	-2.114947
_se	1.514593	.0119436		(Ancillary parameter)		

Obs. summary: 6589.617 left-censored observations at lhrwage<=0
9812.383 uncensored observations

4. *Female Tobit Regression*

Number of obs = 20510
chi2(12) =4077.47
Prob > chi2 = 0.0000
Log Likelihood = -19848.386 Pseudo R2 = 0.0931

lhrwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.2440766	.0055579	43.915	0.000	.2331826	.2549706
exper	.1657891	.0054968	30.161	0.000	.1550148	.1765633
expersq	-.0024626	.0000996	-24.720	0.000	-.0026578	-.0022673
RURAL	-.5598597	.0435123	-12.867	0.000	-.6451473	-.4745722
PROV2	-.5659905	.113057	-5.006	0.000	-.7875911	-.3443898
PROV3	-.4156778	.231145	-1.798	0.072	-.8687404	.0373849
PROV4	-.314898	.1214111	-2.594	0.010	-.5528736	-.0769225
PROV5	-.1673254	.1102729	-1.517	0.129	-.383469	.0488181
PROV6	-.1970929	.1178062	-1.673	0.094	-.4280024	.0338165
PROV7	.2430688	.1087894	2.234	0.025	.029833	.4563046
PROV8	-.1369364	.122142	-1.121	0.262	-.3763444	.1024715
PROV9	-.2659567	.1185571	-2.243	0.025	-.498338	-.0335753
_cons	-4.247019	.1389446	-30.566	0.000	-4.519362	-3.974677
_se	1.990516	.020501		(Ancillary parameter)		

Obs. summary: 14141.67 left-censored observations at lhrwage<=0
6368.328 uncensored observations

5. *Male Two-Step Sample Selection Regression*

```
                                                Number of obs     = 5715065
                                                Model chi2(20)    =87688.55
                                                Prob > chi2       =  0.0000
Log Likelihood =-7384323.4976082
------------------------------------------------------------------------------
     lhrwage |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lhrwage      |
        educ |   .0904996   .0001043    867.437   0.000     .0902951    .0907041
       exper |  -.0060929   .0001091    -55.847   0.000    -.0063067   -.005879
     expersq |   .0003463   1.91e-06    181.397   0.000     .0003426    .0003501
       RURAL |   -.338957   .0007905   -428.794   0.000    -.3405063   -.3374077
       PROV2 |   .0712866   .0015942     44.716   0.000      .068162    .0744112
       PROV3 |   .0005473   .0029363      0.186   0.852    -.0052077    .0063022
       PROV4 |  -.3183571   .0016128   -197.390   0.000    -.3215182   -.3151961
       PROV5 |   .1965451   .0014608    134.549   0.000     .1936821    .1994082
       PROV6 |   .0822933   .0015334     53.669   0.000     .0792879    .0852986
       PROV7 |   .2049308   .0014031    146.052   0.000     .2021807    .2076809
       PROV8 |   .1082069   .0016234     66.656   0.000     .1050252    .1113886
       PROV9 |   .4474011   .0016864    265.295   0.000     .4440957    .4507064
       _cons |   1.419141   .0024139    587.911   0.000      1.41441    1.423873
-------------+----------------------------------------------------------------
probit       |
        educ |   .0360689   .0001394    258.665   0.000     .0357956    .0363422
         age |   .1910848   .0003033    629.920   0.000     .1904902    .1916793
       agesq |  -.0023454   3.78e-06   -619.951   0.000    -.0023528   -.0023379
       RURAL |   -.142627   .0011748   -121.410   0.000    -.1449295   -.1403246
     MARRIED |   .5699008   .0012884    442.334   0.000     .5673756    .572426
     child6 |   .0622616   .0004414    141.059   0.000     .0613965    .0631267
      wnwage |  -.0001533   2.33e-06    -65.896   0.000    -.0001579   -.0001488
       _cons |  -3.697192   .0058826   -628.495   0.000    -3.708722   -3.685663
-------------+----------------------------------------------------------------
         rho |   -0.84702                [_athrho]_cons  = atanh(rho)
       sigma |   .68065529               [_lnsigma]_cons = ln(sigma)
      lambda |  -.57653187   .0006152
------------------------------------------------------------------------------
```

6. *Female Two-Step Sample Selection Regression*

```
                                                Number of obs     = 6239785
                                                Model chi2(20)    =916855.79
                                                Prob > chi2       =  0.0000
Log Likelihood =-5533157.4021643
------------------------------------------------------------------------------
     lhrwage |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lhrwage      |
        educ |   .1167936   .0001844    633.442   0.000     .1164322     .117155
       exper |  -.0061525   .0001895    -32.472   0.000    -.0065238   -.0057811
     expersq |   .0003399   3.44e-06     98.936   0.000     .0003332    .0003466
       RURAL |  -.0871953   .0012975    -67.202   0.000    -.0897384   -.0846523
       PROV2 |   .0339418   .0030208     11.236   0.000     .0280212    .0398625
       PROV3 |  -.0812914    .006585    -12.345   0.000    -.0941977   -.0683851
       PROV4 |  -.3207252   .0032316    -99.246   0.000    -.3270591   -.3143914
       PROV5 |   .0604303   .0029177     20.712   0.000     .0547117    .0661489
       PROV6 |   .0305043   .0031466      9.694   0.000      .024337    .0366715
       PROV7 |   .1814346    .00285     63.661   0.000     .1758487    .1870206
       PROV8 |   .0533501   .0033276     16.033   0.000     .0468281     .059872
       PROV9 |   .3468069   .0032227    107.613   0.000     .3404905    .3531233
       _cons |   .8921527   .0050503    176.654   0.000     .8822543     .902051
-------------+----------------------------------------------------------------
probit       |
        educ |   .0750334   .0001477    508.094   0.000     .0747439    .0753228
         age |   .1946913   .0003425    568.386   0.000       .19402    .1953627
       agesq |  -.0022921   4.29e-06   -534.612   0.000    -.0023005   -.0022837
       RURAL |  -.3757366   .0011623   -323.274   0.000    -.3780147   -.3734586
     MARRIED |  -.1146067   .0012102    -94.704   0.000    -.1169785   -.1122348
     child6 |  -.1378242   .0008725   -157.967   0.000    -.1395343   -.1361142
      wnwage |  -.0000206   3.14e-06     -6.563   0.000    -.0000267   -.0000144
       _cons |  -4.510455   .0066843   -674.787   0.000    -4.523556   -4.497354
-------------+----------------------------------------------------------------
         rho |   -0.39494                [_athrho]_cons  = atanh(rho)
       sigma |    .7240495               [_lnsigma]_cons = ln(sigma)
      lambda |  -.28595776    .001524
------------------------------------------------------------------------------
```

7. *Male FIML Sample Selection Regression*

```
                                                Number of obs     = 5715065
                                                Model chi2(20)    =1030782.7
                                                Prob > chi2       =  0.0000
Log Likelihood =-6912776.4206904
------------------------------------------------------------------------------
     lhrwage |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lhrwage      |
        educ |   .0972178   .0001244    781.794   0.000     .0969741    .0974615
       exper |   .0061507   .0001472     41.774   0.000     .0058621    .0064393
     expersq |    .000155   2.56e-06     60.469   0.000      .00015     .00016
       RURAL |  -.3562123   .0009396   -379.106   0.000    -.3580539   -.3543707
       PROV2 |    .084198   .0021091     39.921   0.000     .0800643    .0883318
       PROV3 |  -.0063897   .0039745     -1.608   0.108    -.0141796    .0014002
       PROV4 |  -.3019047   .0021387   -141.160   0.000    -.3060966   -.2977128
       PROV5 |   .2048627   .0019414    105.526   0.000     .2010577    .2086677
       PROV6 |    .088379   .0020369     43.388   0.000     .0843867    .0923714
       PROV7 |   .2055754   .0018543    110.867   0.000     .2019412    .2092097
       PROV8 |   .1274018    .002152     59.202   0.000      .123184    .1316196
       PROV9 |    .460256   .0022346    205.966   0.000     .4558762    .4646358
       _cons |   1.093439    .003302    331.149   0.000     1.086968    1.099911
-------------+----------------------------------------------------------------
probit       |
        educ |   .0307011   .0001471    208.656   0.000     .0304127    .0309895
         age |   .1926999   .0003269    589.563   0.000     .1920593    .1933405
       agesq |  -.0023799   4.06e-06   -585.852   0.000    -.0023879   -.0023719
       RURAL |  -.1228288   .0011913   -103.103   0.000    -.1251637   -.1204938
     MARRIED |   .5864153   .0015008    390.737   0.000     .5834738    .5893568
     child16 |   .0726316   .0005208    139.466   0.000     .0716109    .0736524
      wnwage |   .0002269   4.22e-06     53.790   0.000     .0002186    .0002351
       _cons |  -3.722272   .0063094   -589.954   0.000    -3.734639   -3.709906
-------------+----------------------------------------------------------------
         rho |   -0.50362                [_athrho]_cons  = atanh(rho)
       sigma |   .73442051               [_lnsigma]_cons = ln(sigma)
      lambda |  -.36987023   .0012115
------------------------------------------------------------------------------
```

8. *Female FIML Sample Selection Regression*

```
                                                Number of obs     = 6239785
                                                Model chi2(20)    =935242.35
                                                Prob > chi2       =  0.0000
Log Likelihood =-5523964.1220517
------------------------------------------------------------------------------
     lhrwage |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lhrwage      |
        educ |   .1241421   .0002478    500.889   0.000     .1236564    .1246279
       exper |   .0002082   .0002412      0.863   0.388    -.0002646     .000681
     expersq |   .0002405   4.21e-06     57.131   0.000     .0002322    .0002488
       RURAL |  -.1195959   .0015075    -79.335   0.000    -.1225505   -.1166413
       PROV2 |    .034688   .0031941     10.860   0.000     .0284276    .0409485
       PROV3 |  -.0804178   .0069743    -11.531   0.000    -.0940873   -.0667483
       PROV4 |  -.3209727   .0034174    -93.924   0.000    -.3276706   -.3142748
       PROV5 |   .0621106   .0030849     20.134   0.000     .0560642    .0681569
       PROV6 |   .0319528   .0033278      9.602   0.000     .0254305    .0384752
       PROV7 |   .1816897   .0030128     60.305   0.000     .1757846    .1875947
       PROV8 |   .0569946   .0035222     16.181   0.000     .0500912     .063898
       PROV9 |   .3508484     .00341    102.889   0.000     .3441649    .3575318
       _cons |    .648085   .0073755     87.871   0.000     .6336294    .6625406
-------------+----------------------------------------------------------------
probit       |
        educ |   .0734357   .0001497    490.624   0.000     .0731423    .0737291
         age |   .1972784   .0003459    570.309   0.000     .1966005    .1979564
       agesq |  -.0023265   4.33e-06   -537.121   0.000     -.002335    -.002318
       RURAL |  -.3741508   .0011625   -321.847   0.000    -.3764293   -.3718723
     MARRIED |  -.1063742   .0012347    -86.156   0.000    -.1087941   -.1039543
      child6 |  -.1306821   .0008899   -146.852   0.000    -.1324262   -.1289379
      wnwage |   .0000485   3.27e-06     14.848   0.000     .0000421    .000055
       _cons |  -4.556076   .0067461   -675.365   0.000    -4.569298   -4.542854
-------------+----------------------------------------------------------------
         rho |   -0.23232                [_athrho]_cons  = atanh(rho)
       sigma |    .73824517              [_lnsigma]_cons = ln(sigma)
      lambda |  -.17150672   .0028578
------------------------------------------------------------------------------
```