DPRU

WORKING PAPERS

Modelling Vulnerability and Low Earnings in the South African Labour Market

Haroon Bhorat and Murray Leibbrandt

No 99/32
ISBN: 0-7992-1977-0

October 1999

Development Policy Research Unit
University of Cape Town

Abstract

Drawing on a comparative overview of the earnings function work on South Africa, this article presents an alternative and comprehensive model of earnings in the South African labour market. The paper uses the standard Heckman two-step approach in trying to model participation, employment and earnings. The results of the modelling indicate, amongst other outcomes, that structural, involuntary unemployment is a crucial feature of the labour market. In addition, rural labour markets and rural participants are seen to be very distinct from those in urban areas – suggesting that spatial rigidities are essential to understanding employment creation in the domestic economy. Finally, age and education are prominent in all the equations of the youth in the labour market, either as non-participants or as the unemployed. The education coefficients show that while non-tertiary schooling encourages participation, it does not secure employment. In turn though, for those in employment, the positive returns to schooling are still reflected.

Acknowledgements

This paper is part of a larger one-year project on Labour Markets, Poverty and Inequality in South Africa. We would like to thank the African Economic Research Consortium (AERC) and the South African Department of Labour for their generous financial assistance.

Table of Contents

List of Tables

1. Introduction

Much work has been done on identifying the correlates of vulnerability in the South African labour market. Most recently, Bhorat and Leibbrandt (1999) provide a descriptive overview of low and zero earners in the labour market, utilising the methodologies found primarily in household poverty studies. One of the key results of the study was that in terms of the race and gender covariates, Africans and females were particularly disadvantaged in the labour market. In addition the study highlighted the importance of rural versus urban labour markets in explaining access to employment and the quality of employment. The importance of education was powerfully displayed through the use of cumulative distribution functions. It appeared though that education was more important in determining earnings than whether an individual gets a job or not. Herein lies the limitation of this descriptive approach: that it is incapable of comprehensively and simultaneously highlighting the different determinants and factors impinging on labour market selection and earnings processes. The next step in such an analysis therefore is to combine these differing covariates, which we identify as important, into an econometric model. Such a model would determine the relative importance of these covariates in explaining each stage of the labour market process, namely participation, employment and earnings.

The modelling work in this paper therefore flows on directly from the descriptive discussion of Bhorat and Leibbrandt (1999) in the sense that we use this analysis to formulate and specify our modelling work. Given the quantum of previous studies on modelling earnings in South Africa however, it is also useful to anchor our approach relative to this recent econometric work. Hence the intention of this paper is two-fold. Firstly, a comparative analysis is undertaken of all the South African earnings function literature, with a focus on the specification of the models and their differing treatments of sample selection issues. Secondly, we propose a model of our own, that attempts to highlight the full dimensions of vulnerability in the South African labour market.

2. Past Earnings Function Models in South Africa

The 1990s have seen the production of a wealth of earnings function work (Moll (1998), Mwabu and Schultz (1996a, 1996b and 1997), Fallon and Lucas (1998), Winter (1998), Hofmeyr (1998), Lucas and Hofmeyr (1998)). This new literature has been spawned largely as a result of the fact that a number of reliable national sample surveys have been conducted in the 1990s. The availability of these data sets have encouraged the application of rigorous and econometrically sophisticated analysis of South African labour market issues for the first time.

We select four of these studies from this literature for further discussion as this is adequate to illustrate the type of choices that need to be made when modelling the South African labour market. It also allows us to illustrate how our approach compares to the existing literature. We summarise the methodology and the results of these studies in Box 1 below.

To the uninitiated, it is hard to read across this literature and make comparisons. The major reason for this is the bewildering array of differences in specification, conceptualisation, estimation techniques and data. These differences are rarely discussed or justified. The four studies presented in Box 1 all use ordinary least squares estimation techniques in estimating the earnings function and all but one use the 1993 South African Living Standards Measurement Survey (LSMS) data. Thus, to a large measure we control for the differences due to data and techniques. This allows us to focus on issues relating to specification and conceptualisation.

In terms of specification issues, each study makes different choices about whether to deal with race, gender and location via dummy variables or via separate equations. Then there are differences in how education, age and experience effects are captured. Some studies use a set of dummy variables and interactive dummy variables for all of these explanatory variables. On the other hand education effects are often assessed through the use of splines.

Box 1: Recent Econometric Approaches to Earnings in the South African Labour Market

Study and Data	Specification of Earnings Function	Coverage of Earnings Function	Labour Market Sample Selection	Results
Mwabu & Schultz (1996a) (*LSMS 1993*)	• Separate earnings equation by race gender and location (Tests against combined model for all) • Three education splines and also includes some tertiary training dummies. • Earnings normalised to wage rate per hour	• Formal sector earnings	• Potential labour market participants (the economically active population) • Sample selection term has to cope with the participation decision and then whether employed, unemployed or informally employed.	• Wage rates of Whites 5 time that of Africans. Half due to education differences • African rates of return to education higher for secondary and tertiary.
Fallon & Lucas (1998) (*LSMS 1993*)	• Separate earnings equations for African, White and Other • Gender and regions specified as dummy variables • Education measured as years. Includes an experience-education interaction effect • Dummy variables for part-time and for public sector	• Formal sector, casual and self-employed earnings	• Actual labour market participants • Sample selection term only has to cope with the unemployed	• Average African gender wage differential is 78% • Average African union wage differential is 71% • Average African public sector premium is 47%
Winter (1997) (*OHS 1994*)	• Separate earnings equations by race and gender. • Includes hours of work as an explanatory variable • Education measured as years	• Formal sector workers	• No sample selection term despite the fullest analysis of participation	• South African participation rates are lower than international trends • Sharp gender differences in participation • Discrimination accounts for over 70% of the gender wage gap for all groups except Coloureds
Hofmeyr (1998) (*LSMS 1993*)	• Separate earnings equations for all earnings categories • Only male earners included • Race and location specified as dummy variables • Education measured as dummy variables • Includes hours of work as an explanatory variable	• Paid employed, including informal self-employed, casual employed, formal non-unionised and formal unionised	• Potential labour market participants • Sample selection term has to cope with unemployed and the unpaid and each of the four paid earnings categories using multinomial logit	• Substantial earnings differentials between the imposed segments of the labour market including formal unionised and non-unionised
***This Study* (*OHS 1995*)**	• *Only African individuals* • *Separate models for male and female* • *Tests against combined model for urban and rural* • *Education splines*	• *All employees and self-employed*	• *Potential labour market participants* • *Selection into the labour market and then into employed*	

For us, these choices are largely defined by the descriptive analysis in Bhorat and Leibbrandt (1999), as alluded to above. This picture revealed that the vulnerable are almost exclusively found within the African and Coloured racial groups, with the African group accounting for close to 90 per cent of all low earners and no-earners. In our modelling we therefore confine our attention exclusively to African individuals. We also know that, within the African group, females carry a larger than proportionate burden of low participation, high unemployment and low earnings. In addition, low-earning African females tend to be found in different sections of the labour market to males. Given these factors, there is a strong likelihood that estimates based on aggregate African models are likely to throw up average parameters that are not useful representations of either male or female groups. In addition, we explicitly want to compare African female and male models. Thus, in all instances we run separate estimations for African males and females.

Thus, on the basis of this descriptive support, we are prepared to impose these restrictions. Such restrictions are also in line with the more careful econometric work represented in Box 1. We are confident, therefore, that they will improve the quality and usefulness of the resultant estimates.

Besides these racial and gender dimensions of vulnerability, the descriptive analysis also revealed strong rural and urban differences within both African male and female groups. It is important for our policy conclusions that we explore these differences. Initially, we do so by estimating models for all African women and all African men in which we include a rural and urban dummy variable. We then go on to estimate separate models for rural and urban areas so that we can compare the coefficients and statistically test for significance between these coefficients. Thus, in assessing rural and urban differences, we do not impose separate specifications from the outset. Rather we assess the specifications as part of the estimation process.

These are the major choices that we make regarding our earnings equations. As we explain later, we estimate labour participation and employment equations alone with our earnings equation. Each of these three equations include certain explanatory variables that clearly pertain to that equation and not to the others. However, our earlier descriptive analysis makes it clear that there are age, education and provincial aspects to labour market vulnerability at each of these three levels. All three equations will include a set of dummy variables capturing age and provincial effects and a set of three educational splines that capture the returns to schooling at primary, secondary and tertiary level.

On the conceptual level, hardly any of the South African work spells out even a rudimentary model of the South African labour market as the context for estimation. Earnings function work only makes sense against such a context and part of the difference between the models must lie in the fact that the earnings functions are set up, often only implicitly, in differently defined labour market contexts. We tease out this point by a close examination of the sample selection equations that are used in each of the studies. Each researcher chooses a sample selection equation based on a demarcation of the relevant sample (labour market) of the study as well as the relationship between the sub-sample of earners and this broader sample. Thus, inspection of the interface between the earnings equation and the sample selection equation reveals much about the overall labour market context within which the earnings function work is located.

We illustrate with reference to the four studies presented in Box 1. The key columns are the two columns reflecting the coverage of the earnings function and labour market sample selection.

The Mwabu and Schultz (1996) study is the most careful of all four studies in terms of testing for the adequacy of different specifications for the earnings function. However, the focus of the earnings function, formal sector earnings, is assumed from the outset and not derived. The selection equation begins with all potential labour market participants. It includes an extensive array of agricultural asset variables that are the hallmarks of a participation equation in a conventional developing country. However, the resultant selection term is insignificant in all but one of the earnings functions and it is therefore omitted for the final set of earnings function estimations. Indeed, as Mwabu and Schultz point out, these variables are jointly insignificant even in the participation equation thus raising some problems for the identification of the two-equation model.

To us this insignificance is hardly surprising for two reasons. First, one of the apartheid legacies in South Africa is the decimation of any small-holder and subsistence farming classes (Lipton et al, 1996). Thus, it is hard to conceptualise any clear relationship between these agricultural assets and labour market participation. Second, the earnings equation is narrowly focused on formal sector earnings. This leaves participation in the labour market, selection into employment and participation in the informal sector to be dealt with by the participation equation. We would expect such a diversity of forms of participation and selections to be inadequately captured by a single participation equation. Even assuming that all unemployment in South Africa is voluntary, and therefore indistinguishable from the decision regarding whether or not to participate in

the labour market, the participation equation also has to deal with the awkward issue of participation in the informal sector versus the formal sector.

The Fallon and Lucas (1997) study covers a far broader section of the labour market in the earnings function itself. Formal sector employees, the self-employed and part-time workers are all included as earners. The selection equation then selects from the chosen sample of all labour market participants into this reduced sample of earners. The selection equation therefore covers the selection from a the pool of participants into earnings; i.e., an employment-unemployment equation. Of course this makes the selection equation coherent and interesting in its own right. However, this coherence is achieved at the cost of ignoring the issue of participation in the labour market and therefore using a narrower sample than the other studies. The employment probit includes a set of variables defining "other household income". These variables would usually be thought of as factors influencing participation rather than factors influencing employment. The exception would be if unemployment were viewed as voluntary. Fallon and Lucas clearly do not believe this to be the case. However, this then leaves the participation-unemployment nexus hanging in the air in this study.

Winter (1997) offers a full analysis of participation in the South African labour market. Indeed it was her clear documentation of the importance of South Africa's very low participation rates and the gender and racial biases in these participation rates that informed our insistence in this study that participation is one of aspects of labour market vulnerability in South Africa. Having provided this exhaustive analysis of participation, Winter uses her earnings function work to document the importance of earnings discrimination by gender in the South African labour market. The focus of this earnings analysis is on formal sector workers. In estimating earnings functions by gender she does not include a sample selection term. Indeed she could not as she has provided extensive coverage of participation but no coverage of unemployment. She has left the selection into employment unexplored and therefore has a missing sub-sample in her labour market.

Like Fallon and Lucas, Hofmeyr (1999) attempts to capture all earners within the ambit of the earnings function estimations. Hofmeyr uses the same earnings categories as Fallon and Lucas but goes further by splitting formal sector workers into unionised and non-unionised sections. However, Hofmeyr differs from all previous studies in his approach to selection. He sets up a full sample of potential labour market participants and presumes that they are allocated into one of his four categories of earners or into unpaid household help (helping another household member who is self-employed) or into no employment. This selection is done simultaneously in a multinomial logit allocation equation in which “no employment” is defined as the default category. It is interesting to see how the characteristics of those allocated into the earnings segments differ from those without employment. However, it needs to be stressed that “no employment” covers non-participants and unemployed. Thus, the model can not provide useful information on either participation or on unemployment.

The original rationale for such a multinomial logit model is an occupational choice model (Roy, 1951). Hofmeyr is well aware of the fact that the South African labour market offers an uncomfortable context for such a choice-theoretic view of the allocation process and wants the model to cover both supply and demand elements and therefore choice and constraints from the individual point of view. It is not clear that the model is up to such a task as is evidenced by the fact that the model allocates many individuals to incorrect segments of the labour market.

Hopefully this review of four recent econometric studies has provided a relevant and useful context for the presentation of our approach to modelling. Our special focus is on the vulnerable in the labour market. Previous empirical work has made it quite clear that vulnerability needs to be defined in such a way that it encompasses labour market participation, selection into employment as well as the determinants of earnings. The biggest conceptual issue that we face with regard to the formulation of our modelling is to give detailed attention to all three of these stages in the labour market.

3. The Model Set-up

Our model structure deals with these stages sequentially. First, we begin with a full sample of potential labour market participants and estimate a participation probability model. Then, for the reduced sample of labour market participants we estimate an employment probability model. Finally, we estimate an earnings function using the sample of employed Africans. Such a sequential model can be loosely justified by the assumption that labour market participation and employment are first choice activities of all potential labour market participants and we are therefore modelling a rationing process. The participation equation attempts to throw light on the key factors selecting participants. Once the participants are determined, the second stage models the employment allocation process. The final stage models earnings of those who succeed in obtaining employment.

This is certainly a plausible South African scenario; particularly for the employment-unemployment step between participation and earnings. We argued above that other econometric studies of the South African labour market have tended to blur the distinction between participation and unemployment in their selection equation. While this is not particularly important if the purpose of the exercise is to cleanse the earnings equation of sample selection problems, it is of no use if the purpose of the analysis is to examine the determinants of participation and employment.

Such analysis is particularly important in the South African context because of the debates that exist over usage of the narrow versus the expanded definition of unemployment (ILO, 1996 and Nattrass and Seekings, 1998).[1] In discussions over the two unemployment definitions, insufficient attention has been given to the fact that a movement from a broad to a narrow definition of unemployment involves an assertion that discouraged workers are not participating in the labour force.[2] Thus, the sub-sample of unemployed shrinks to the narrow definition and the sub-sample of participants expands to take in the discouraged work-seekers. By distinguishing between participation and unemployment we can assess the difference that the change in definition makes to participation and unemployment.

Related to the narrow versus broad unemployment issue is the question of voluntary versus involuntary unemployment. All analysts recognise that unemployment is predominantly involuntary in South Africa. Even more important is the fact that the unemployment questions in all recent surveys are designed to select out those who want jobs but do not have them from the sample of potential labour market participants. Thus, the surveys themselves are structured to capture the involuntarily unemployed. Yet, as pointed out earlier, the earnings function literature in South Africa has tended to present a messy interface between participation and unemployment in their selection equations. Indeed, given that most selection equations are starkly framed in terms of participation versus non-participation in the labour market, it is only by assuming that unemployment is voluntary that the specified selection equations can be made tenable. By including both participation and employment equations in our work, we are clearly defining unemployment as a state that occurs despite a decision to participate in the labour market. It is therefore clearly involuntary.

Our estimation starts out with a full sample of potential labour market participants. It then shrinks the sample to cover actual labour market participants and then shrinks the sample further to cover earners. It is now well established in the labour economics literature that the estimates derived in the employment model and in the earnings model may be biased because of the fact that they are both based on non-random, reduced versions of the original sample of potentially employable Africans (Heckman, 1979). Thus in all versions of our modelling we control for the possibility of sample selection problems. We use a probit model to estimate our participation equation. Then we use another probit model to derive employment probability estimates conditional on the characteristics of all labour market participants and *conditional on the fact that these are the actual participants taken from a full sample of all potential participants*. Then we derive estimated earnings coefficients conditional on the individual characteristics of the earners and *conditional on the fact that*

[1] The formal distinction between these two categories is extensively discussed in Bhorat & Leibbrandt (1999).

[2] The ILO (1996) argues that there are so many discouraged workers that they must be doing something. In other words, the discouraged worker category is an artifact of inaccurate survey work. This is a plausible argument for some surveys data sets. However, as we argue in Bhorat & Leibbrandt (1999) and in Bhorat (1999), the OHS 95 gives serious attention to these issues and we would therefore argue that the patterns are robust enough to accept. It is interesting to restate the central conclusion of our earlier review of the unemployment issue. In OHS 1995 the discouraged worker category is notably smaller than previous estimates but the narrow unemployment category is larger. This suggests that part of the inaccuracy of earlier survey work may have involved an inaccurate capturing of search activity.

these earners are a sub-sample of all labour market participants and an even smaller sub-sample of potential participants.

In each instance we use the Heckman two-step approach to cope with the sample selection issue (Greene, 1993 and Breen, 1996). Having estimated the participation probit, we use these estimates to derive our estimate for the inverse Mills ratio (lambda) for inclusion in employment probit. It is the inclusion of this lambda that allows us to make the employment probit conditional on positive participation. We then use the estimates from the employment probit to derive a new estimated Mills ratio, reflecting selection into earnings. The inclusion of this second lambda in the earnings equation makes the earning equation conditional on participation and selection into employment. It seems plausible to argue the selection into employment and the determination of earnings for those employed are simultaneous processes rather than sequential. We also allow for this possibility by deriving another set of estimates for the employment probability model and the earnings function based on a single, integrated maximum-likelihood model.

One of the strengths of a clear delimitation of participation, employment and earnings stages in the labour market is that it facilitates the selection of a coherent set of variables for each equation. For example, as mentioned when we discussed Fallon and Lucas (1997) above, it is fairly common to see household variables in an employment-unemployment probit. However, such variables would normally relate to a participation process rather than an employment process. Thus, our participation equation includes a full set of household composition variables by age as well as variable reflecting income from other household members (and the square of this variable to allow for non-linearities). In terms of the two-stage selection model, these household variables identify the lambda that is included in the employment probit.

The employment equation therefore only contains information about the personal characteristic of each job seeker (age, education and location). As these variables are all also plausible explanatory factors in the earnings function, this raises a tricky identification issue in terms of the selection lambda that is derived from the employment probit for inclusion in the earnings equation. There are two factors that lead us to suspect that this is not a problem in our estimations. First, age would seem to be important in the employment/unemployment equation whereas potential experience (and potential experience squared) would appear to be the more relevant age-related variable for the earnings function. Thus, age effects are specified differently in the two equations. Second, the lambda carried through into the earnings equation incorporates the first lambda from the participation equation as an identifying explanatory variable. This lambda is an additional variable in the employment equation.

3.1. Data Issues

Thus, there seems to be a comforting degree of agreement between tidy econometric practice and the type of labour market that we estimate in order to capture the key aspects of labour market vulnerability in South Africa. However, it would be disingenuous of us not to conclude this section by clearly spelling out the constraints that the data have imposed on our modelling. One key limitation is the inability to use the survey to clearly demarcate an informal and a formal sector. Models of segmentation in developing countries give explicit attention to these earnings segmentations (Glick and Sahn, 1997, Heckman and Hotz, 1986 and Andersson, undated). We cannot do this.[3]

Yet, descriptive analysis highlights the fact that, for Africans, self-employment clearly offered inferior earnings. However, further analysis showed that it was African female domestic workers who dominated this self-employment category (Bhorat,1999). As we are estimating separate earnings equations by gender with full sets of sectoral and occupational dummy variables and an explanatory variable for hours worked, this self-employment effect will be adequately captured in the female earnings equations.

Our participation equation is also far from perfect. It is common to define potential labour market participants by age (16-65). This is the definition used in earlier descriptive analyses in this report. However, if we follow through with this definition here, then the non-participant sub-sample is dominated by young adults who are still in education. It might be the case that some young adults are staying in school because of poor employment prospects in the labour market. However, given, high repetition rates and educational backlogs in South Africa, the routine school-leaving age is also well above 16 years. Such people are not potential

[3] Our review of Fallon and Lucas (1997) and Hofmeyr (1998) showed that the LSMS data is similarly flawed when it comes to an analysis of the informal sector. It would appear therefore that there is no data set in South Africa that can be used to explore formal sector/informal sector interactions in South Africa. The problem of uncovering the informal sector in the OHS95 data set is again taken up in Bhorat & Leibbrandt (1999) and Bhorat (1999).

labour market participants. However, not all young adults are in school and this pattern is overlayed with a high youth unemployment problem. Therefore, it would be distortionary to deal with this issue by raising the age of labour market participants. Rather we remove all people who are in education from the sample.

This significantly reduces the sub-sample of non-participants. As the table below illustrates, the number of participants enrolled in education is just over 4 million, accounting for close to one-quarter of all African labour market participants. In our derivation of labour market participants then, we firstly exclude those in the last category, being either retired, permanently disabled or unclassified. The exclusion of those in education, of course omits the largest number of individuals.

Table 1: African Participants, by Type of Activity

Activities	Male	Female	Total
Working Full-time	3597992 42.71	2009485 22.29	5607477 32.15
Working Part-time	245596 2.92	294602 3.27	540198 3.10
With a job, but absent from work	39512 0.47	30360 0.34	69872 0.40
Going to school/university/college	2061942 24.48	2039084 22.62	4101026 23.51
Unemployed (looking for work)	1677274 19.91	1981823 21.98	3659097 20.98
Not working, not looking for work	293627 3.49	678380 7.52	972007 5.57
Housekeeping	21096 0.25	1337700 14.84	1358796 7.79
Retired (pensioner)	253188 3.01	431601 4.79	684789 3.93
Permanently unable to work	233189 2.77	211216 2.34	444405 2.55
Other	1060 0.01	2224 0.02	3284 0.02
Total	8424476 100	9016475 100	17440951 100

For African females there remain a large number of non-participants who are engaged in home production, but very few males in this category. The structure of the survey is such as to classify all male non-labour market activity as self-employment, to impute earnings to these activities and include such males as unregistered self-employed earners. With the removal of African individuals in education, the sub-sample of non-participants as a whole drops to 12 207 447, 70% the size of the original sample. Of this narrower sample then, 11% of the activities undertaken involve home production. Of this 11%, the overwhelming majority (98%) are female.

The import of the above decisions taken in dividing the sample of African participants, for our modelling is as follows:

- When the broad definition of unemployment is adopted, we have a very small sub-sample of male non-participants and the female sub-sample are exclusively those engaged in home production.
- When we adopt the narrow definition of unemployment, male non-participants are dominated by discouraged workers and female non-participants are a mix of discouraged workers and women engaged in home production.

There is one final data difficulty in the participation equation. It is not possible to attribute children to specific parents. We include a variable capturing the number of children in the household but this is certainly only a loose proxy for the influence of own children on participation.

4. Model Results

Tables 2 to 5 present the influence of the different covariates on the probability of participation and employment as well as on the level of earnings of the employed. For the covariates which are dummies, the following are the referent variables:

- Location: Urban
- Age: 16-24
- Province: Western Cape
- Sector: Agriculture
- Occupation: Farm worker
- Union Status: Union member

As explained above, the equations are all run for African individuals only. In addition, separate male and female equations are estimated for both the expanded and strict definitions of unemployment. The key results for participation, employment and earnings, respectively, are presented in tables 2, 3 and 4 below. Tables 5 through 9 in the appendix represent the output when all of these models are re-estimated separately for rural and urban areas.

4.1. Participation Equation

Table 2 presents the results from the participation decision in the labour market. The urban dummy variable is significant for females but not for males, across both the narrow and expanded definitions of unemployment. Hence, for females, living in an urban area increases the probability of participating in the labour market, while for males location has no bearing on their participation decision. Further evidence in this regard comes from tables 5 and 8 in the appendix. These tables present results for male and female participation equations in urban and rural areas. The coefficients in the male equations in both urban and rural areas are very similar to each other and to the coefficients in the aggregate model. It would seem that there are no noteworthy differences in male participation in urban and rural areas. However, this is not always the case with African females and we will flag these differences in our discussion below.

The education splines suggest that schooling is an important variable in determining whether individuals participate or not in the labour market. For African males, according to the expanded definition both primary schooling and secondary schooling have a positive bearing on the participation decision. Surprisingly, having tertiary education does not appear to influence the decision to participate or not. This insignificance could be due to the relatively small share of African males with tertiary education (6.5%), coupled with the fact that this level of education will not determine a decision of whether to enter the labour market or not. However, with a switch to the narrow definition of unemployment and the consequent re-classification of the discouraged workers as non-participants, all three splines become significant. The significant tertiary variable here implies that, tertiary education greatly increases the probability of being employed or of being an active job seeker relative to being one of the discouraged work-seekers who now dominate the non-participants.

For females, the education splines are slightly different. Only secondary education is significant for the expanded definition, while for the narrow definition, secondary and tertiary schooling are significant. As with males, a small percentage of females have tertiary education. Remembering that the non-participants here include discouraged work-seekers, the data shows that of the female non-participants by the narrow definition, only 1% have tertiary education compared to 11% for participants. Again, the possession of secondary or tertiary education does distinguish females who are employed or actively searching for employment from those who do not participate.

Table 2 : African Male and Female Labour Participation Equations for Expanded and Narrow Definitions of Unemployment

	Male				Female			
	Expanded		Narrow		Expanded		Narrow	
	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar
Urban	.0099	.504329	.0072	.504329	.1418*	.390973	.1321*	.390973
None-Std5	.0029*	4.43131	.004**	4.43131	.0027	4.20941	.0030	4.20941
Std 6-10	.0052*	1.44132	.0159*	1.44132	.0517*	1.27643	.0568*	1.27643
Tertiary	.00241	.105839	.0161*	.105839	-.0231	.103659	.0115**	.103659
26-35	.05796*	.349291	.13893*	.349291	.09665*	.342574	.1309*	.342574
36-45	.07255*	.248674	.18949*	.248674	.07898*	.252344	.1831*	.252344
46-55	.05132*	.143162	.1616*	.143162	-.0115	.149113	.1371*	.149113
56-65	.01766*	.053417	.1440*	.053417	-.12882*	.050201	.0781*	.050201
No. of Kids <7	.00166	.785774	.0060	.785774	-.02125*	1.11372	-.0258*	1.11372
No. of Kids 8-15	-.00156	.878397	-.0095*	.878397	-.00989*	1.14248	-.0168*	1.14248
No of males 16-59	-.009189*	1.96432	-.0317*	1.96432	-.00988*	1.33687	-.0164*	1.33687
No of fems 16-59	-.00937*	1.51446	-.0259*	1.51446	.02978*	2.15132	.0116*	2.15132
No of Adults >60	-.03313*	.319228	-.0958*	.319228	-.00307	.34587	-.0381*	.34587
Other hhld income	-6.12e-07*	17352	-9.90e-07*	17352	-2.24e-06*	20880.8	-9.73e-07*	20880.8
Other hhld income sqred	2.75e-12*	1.1e+09	7.16e-12*	1.1e+09	2.44e-12*	2.0e+09	4.83e-13*	2.0e+09
Obs. Prob	**.91173**		**.7753**		**.6584**		**.4865**	
Pred. Prob (at x-bar)	**.9348**		**.80972**		**.6773**		**.4878**	
No Obs	**15658**		**15658**		**19548**		**19548**	
Chi(2)	**1084***		**2450**		**2190**		**2144***	
Pseudo R2	**0.1120**		**0.1426**		**0.0870**		**0.0792**	

* Significant at the 1% level

** Significant at the 5% level

What this suggests is that education is important in determining whether an individual participates or not. However its significance seems to increase, when using the narrow definition of unemployed. This is manifest in much better educational qualifications amongst participants, relative to non-participants. For both females and males, non-participants by the narrow definition are dominated by discouraged work-seekers. This educational wedge between the two cohorts, drives the results in the narrow definition equations.

The age dummy variables are all significant barring the case of females 46-55 under the expanded definition. In addition, all significant coefficients have the same positive sign, barring the case of females 56-65 under the expanded definition. In other words, the age dummies suggest that the probability of participation increases for all age cohorts, relative to the youngest cohort, namely 16-25 years. This is not a surprising result, as those adults who are older are more likely to have a job or seeking a job, irrespective of the definition of unemployment used. However, the fact that this age effect strengthens with a move to the narrow definition of unemployment is alarming as it suggests that there are a significant proportion of the youth cohort that are discouraged worker seekers.

While not presented in Table 2 above, the equation also included a full set of provincial dummy variables. These dummies generally had similar results across the genders and definitions. Provinces with significant results were the Northern Cape, Kwazulu-Natal, North-West, Mpumulanga and the Northern Province. The referent province was the Western Cape. In each of these cases, being in the respective province decreased the probability of participating in the labour market relative to those in the Western Cape. All of these provinces have a higher percentage of rural economically active than the Western Cape. For the narrow definition, these results also pick up the much larger number of discouraged work-seekers in these provinces relative to the Western Cape.

The Household block of variables, includes two 'number of children' variables, three 'number of adults' variables and two household income covariates. Here the gender biases of child rearing become immediately evident. For males, the number of children of any age in a home is insignificant in determining their participation decision. For females, however, it is clearly established that the greater the number of children under the age of 7 or between the ages of 8 and 15, the less the probability of their participation in the labour market.

With the exception of adults older than 60 in the female expanded equation, the 'number of adults' variables are all significant. What is interesting though is that in most cases, the coefficients are negative. This indicates that the presence of a greater number of adults in the household acts as a deterrent to participation in the labour market. For females though, the signs are positive when considering the number of female adults aged 16-59 in the home. In other words, females are more likely to participate in the labour market, the larger the number of working age women in the home by both definitions of unemployment. This fact may be picking up those women involved in home production, who because they will not be participating cause other females to participate in the labour market. The more working age males in the home though, the less likely are women to participate.

While the larger the number of aged in the home causes the probability of participation to fall for males by both definitions, this is not true for females. For females, the expanded definition estimate is insignificant, while the narrow definition is significant. These results in general suggest that for males and females, the presence of an aged person (and in all likelihood a pensioner) acts as a deterrent to participation in the labour market.

Finally, the household income variables are both significant across genders and definitions, with the same negative sign. It is evident that the greater the value of other household income available to an individual, male or female, in a household reduces the probability of their participation in the labour market. In other words, access to income within a household is an important determinant in an individual's decision to participate. However, the small but positive values on the household income squared coefficients suggest that this effect is dampened as income increases.

Tables 5 and 8 in the appendix show that, in a few key areas, the aggregate female participation patterns that we have discussed above have blurred important rural-urban differences. We highlight two cases. First, the education results for urban females are stronger than for the whole sample of females. Thus, for urban women under the expanded definition, only primary schooling is significant in increasing the probability of participation. For the narrow definition, all three educational splines are significant. This would suggest that for urban women, their educational qualifications are a more important determinant of their decision to participate, when compared with the sample of all females.Second, for urban women, the presence of children between the ages of 8 and 15 is not significant in determining participation, across either definitions of

unemployment. This would suggest that in urban labour markets, women are less likely to give up a job or stop searching for a job due to older children being in the home. It may also reflect a work life-cycle phenomenon, where women after rearing the children at home, then re-enter the labour market. Noticeably this is a purely urban characteristic, as this variable is negative and significant for rural females.

4.2. Employment Equation

Having considered the determinants of participation, we retain the sample of those individuals who decide to participate, and in turn estimate the probability these participants have, in finding a job. The results from the employment probit are presented in Table 3 below. Maintaining consistency with the participation models, we also estimated separate employment equations for urban and rural areas. These estimations are reported in tables 6 and 9, respectively, of the appendix. Note that there were too few narrowly unemployed females in urban areas for the urban, female employment equation to generate a set of estimated coefficients.

Many of the variables in the employment equation are the same as those included in the participation equation. However, we do not include household structure or household income variables in the employment equation. As discussed in an earlier section of this paper, the employment equation is set up to capture the rationing process through which jobs are allocated to some of those who are seeking work. The household variables are seen to influence the decision to seek work but not the process finding employment.

We begin with the last variable first. The coefficients for lambda are significant for males and females for the narrow definition, but, under the expanded definition, only for males. Lambda represents the inverse Mill's ratio and is a measure of the selectivity bias in the sample. The significant results suggest that sampling bias did exist in the sample and needed to be corrected for through this procedure. Labour market participants do not look like a random sample chosen from all of the economically active population. This difference is particularly acute when participants are defined based on the narrow definition of unemployment.

The location results show that for African males across both definitions, living in an urban area reduces the probability of being employed. For females the result also holds for the narrow definition of unemployment. Given that employment opportunities present themselves overwhelmingly in urban areas, the negative coefficients are seemingly surprising. Tables 6 and 9 in the appendix allow us to unpack this a little further. These tables contain figures for the actual and estimated probabilities of employment in urban and rural areas, respectively. It can be seen that both of these probabilities are very close for urban and rural areas. The predicted probabilities of employment are based on an average set of characteristics for urban or rural work-seekers, respectively. The mean values for all variables which are shown in tables 6 and 9 show that the average rural work-seeker is not as well educated or as well located as the average urban work seeker. The marginal effect of the urban-rural dummy variable in table 3 is based on an average set of characteristics for the *combined* urban and rural sample. Thus, it assesses the probability of employment for an average worker who has characteristics that lie in between those reflected in the separate urban and rural estimations. This worker has less favourable attributes than the average urban worker and, *for such a person*, rural areas offer a higher probability of employment.

The case of female participants under the expanded definition appears to offer an important exception. Tables 6 and 9 show that the actual and predicted probabilities of being employed are close to ten percent higher in urban areas than rural areas and the urban/rural dummy variable in table 3 is positive reflecting a higher probability of employment in the urban areas. However, even here, the urban/rural dummy variable is not statistically significant. This reflects the fact that an African female with average aggregate characteristics would have better characteristics than the rural average and worse than the urban average. She would therefore have a higher than fifty percent chance of employment in rural areas and a lower than sixty percent chance of employment in urban areas.

Table 3 : African Male and Female Employment Equations for Expanded and Narrow Definitions of Unemployment

	Male				Female			
	Expanded		Narrow		Expanded		Narrow	
	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar
Urban	-.08119*	.518166	-.06044*	.536173	.01502	.470573	-.11547*	.512476
None-Std5	-.01214*	4.46674	-.00877*	4.49146	-.00381**	4.4932	-.00940*	4.58104
Std 6-10	.00911	1.46147	.00007	1.49287	.03556*	1.53057	-.02063*	1.66111
Tertiary	.04735*	.108512	.03623*	.116307	.14162*	.126277	.15301*	.153247
26-35	.00314	.356938	-.00214	.357391	.2068*	.378007	.08557*	.370626
36-45	.05499*	.261804	.01259	.27969	.35464*	.259706	.16435*	.278516
46-55	.11750*	.146699	.04410*	.156069	.38832*	.129088	.2040*	.143411
56-65	.25410*	.049023	.09665*	.053912	.41972*	.033449	.23622*	.039938
E.Cape	-.06235**	.12238	.04477**	.108051	-.12400*	.162958	.05221	.15373
N.Cape	.00402	.01065	-.02210	.01137	-.0151	.007157	.02185	.007811
Free Stat	-.00515	.087579	.07063*	.089931	.03244	.089192	.0729**	.091194
Kwaz/Natl	.03850	.192919	.02811	.192312	-.04242**	.221515	.05947	.225838
North-W	-.01404	.115938	.04595**	.115271	-.04376	.098021	.09928*	.093327
Gauteng	.05342*	.249739	.04712*	.268261	.00861	.193928	.03198	.219473
Mpumal	-.0207	.092567	.06406*	.090007	-.08506*	.0792	.07478**	.071484
N.Prov	.08420*	.08616	.08594*	.079052	-.17369*	.118059	.06257*	.10429
Lamda	-1.4131*	.156201	-.43850*	.329335	.0018	.502512	-.44005*	.740367
Obs. Prob	.7173043		.8434		.5460		.73894	
Pred. Prob at x-bar)	.7419969		.8740		.55423		.76711	
No Obs	14203		11931		12810		9426	
Chi(2)	2677*		1585		1902		1245	
Pseudo R2	0.1483		0.1548		0.1078		0.1156	

* Significant at the 1% level

** Significant at the 5% level

The education splines firstly show that across both genders and definitions, the possession of primary schooling or less reduces the probability of finding employment. Indeed, for females by the narrow definition, this negative coefficient holds for secondary schooling as well. In contrast the coefficient on tertiary education is positive across both genders and definitions. Collectively the education splines indicate that individuals with lower levels of education have less of a chance of getting a job than those with high-level, and specifically tertiary, education. This analysis confirms time-series labour demand analysis done elsewhere on the South African labour market. Such studies indicate that labour demand patterns reflect a growing demand for higher skilled labour, and stagnant or declining demand for less skilled workers (Bhorat and Hodge, 1998).

The age variables, as with the previous equation, are not surprising as they show an increased probability of employment in older age cohorts relative to those in the 16-25 group. This reflects the large number of youth who are unemployed. The insignificant results for all except one age cohort for males by the narrow definition, may be picking up the large number of discouraged work-seekers who are fairly evenly distributed across these age groups. The provincial results are mixed. Some of the provinces, such as the Northern Cape and Kwazulu-Natal yield mostly insignificant results. For the rest, in some cases there is a lower probability of being employed in the province, relative to the Western Cape, and in others a higher probability. In Gauteng for example, African males have a greater probability of finding employment than their counterparts in the Western Cape. The parallel coefficients for females though, are insignificant. The Northern Province, one of the poorest provinces in the country, yields positive coefficient except for females by the expanded definition. A factor that may be influencing these results, is the large Coloured labour force in the Western Cape, that means a much lower share of African employment in the province relative to the rest of the country. Indeed, the Western Cape, while accounting for 14 % of total employment in the country, only accounts for 3% of African employment.

We have already referred to the separate urban and rural employment estimations that are presented in tables 6 and 9 of the appendix. We conclude this section by noting further interesting results from these tables. For males for example, secondary education is seen to be important in predicting employment in urban labour markets. The insignificance of secondary education in the aggregate male employment equations therefore reflects the lack of significance of secondary education in rural areas. Contrary to these mixed results for secondary schooling, across all four equations in both urban and rural areas, tertiary education is crucial in predicting employment. Noticeably, the effect of primary schooling or less is weaker in rural areas. The location cuts also show more consistent results for the provincial dummies. Along with Gauteng, the Western Cape is seen to be the most favourable location for rural work-seekers. However, this is not as clear cut for urban work-seekers especially when discouraged unemployed are not included as labour market participants.

4.3. The Earnings Function

Table 4 presents the earnings function for all those employed, by gender and again by the two definitions of unemployment. The move from narrow to expanded unemployment does not affect the classification of earners but only affects the sample selection variable (lambda) in the earnings function. Thus, the results of the estimations do not and would not be expected to differ much by the choice of narrow versus expanded unemployment. However, as employment and earnings were estimated together in one maximum-likelihood process, we continue to report the two sets of earnings estimates. Once again we report disaggregated urban-rural equations in table 7 and 10 of the appendix. In all estimations earnings are measured by the log of the monthly total wage earned by individuals, which is the manner in which the survey reported total pay.

From the results it is clear that being in an urban area increases the earnings of the employed. It is an effect that holds true for males and females and for both definitions of unemployment. The education splines are particularly interesting. They show that for African males and females primary schooling or less as well secondary schooling is important in increasing earnings, but not tertiary education. Tables 7 and 10 show that the insignificant impact of tertiary education is true for the disaggregated urban and rural estimates as well. Hence, while tertiary education has been shown to be crucial in determining whether an African individual gains employment, it is not relevant in predicting the level of earnings. Notice that the rates of return to secondary schooling are in each case higher than the returns to primary schooling or less. Hence the return to earnings of one additional year of secondary schooling range from 8.1% to 10.9%, while in the primary

schooling case, the figures are 3.5% and 5.1%. Furthermore, the returns to males on secondary education are higher than for females, but lower than females in the case of primary education. Males also get higher returns to education in urban areas than in rural areas but the returns to females do not appear to differ in this way.

Table 4: African Male and Female Earnings Equations for Expanded and Narrow Definitions of Unemployment

	Male		Female	
	Expanded Unemployment	**Narrow Unemployment**	**Expanded Unemployment**	**Narrow Unemployment**
Urban	.1192892*	.1294798*	.1780*	.1912*
None-Std5	.034631*	.0357029*	.0488*	.0514*
Std 6-10	.1087725*	.1078169*	.0816*	.0927*
Tertiary	.0367241	.0312919	.0234	.0318
E.Cape	-.1070717*	-.1116288*	-.1117**	-.1465*
N.Cape	-.155832*	-.147461*	-.2329*	-.2267*
Free Stat	-.3030291*	-.3149274*	-.5577*	-.5694*
Kwaz/Natl	.0350547	.0377735	.0538	.0411
North-W	-.0153048	-.0194882	-.0891	-.1180**
Gauteng	.0536299	.052913	.1422**	.1357**
Mpumal	-.0353655	-.0467185	.1254**	.0945
N.Prov	.1192166*	.1087125*	.2051*	.1613*
Mining	.607814*	.6068056*	.2840**	.2831**
Manuf	.6394293*	.643446*	.2494*	.2535*
Electricity	.8829402*	.8850715*	.5171*	.5237*
Constr	.4777885*	.4826215*	.3753*	.3761*
Wholes	.5040102*	.5100534*	.1957**	.1995*
Transport	.7384904*	.7407738*	.5106*	.5179*
Finance	.6486269*	.6500433*	.4674*	.4708*
Comm Serv	.677428*	.6803405*	.3619*	.3653*
Other	-.2943609*	-.291168*	.2713*	.2735*
Armed Forces	.5296329*	.5256727*	.7866	.7881
Managers	.7602302*	.767167*	.9501*	.9552*
Profess	.7286019*	.7230151*	1.029*	1.031*
Technicians	.4671531*	.4656222*	.9203*	.9212*
Clerks	.2231234*	.2237144*	.5926*	.5953*
Serv&Sales	.1635076*	.163773*	.3514*	.3532*
Skilld Agric	.1874371*	.1916955*	.0711	.0797
Craft	.1862878*	.1865975*	.2341*	.2371*
Mach Operator	.1460807*	.1461151*	.3355*	.3383*
Unspecif	-.0151913	-.0098603	.1577	.1641
Domes Helper	-.0466067*	-.0482176	.2067*	.2073*
Mining lab	-.0668179**	-.0656155	.2000	.2033
Manuf lab	-.0591869*	-.0579945*	.2727*	.2732*
Trprt lab	-.0401749	-.0409864	-.2676	-.2347
Domes Worker	-.8043337*	-.7996188*	-.3591*	-.3596*
Union Member	.1997917*	.1941152*	.2131*	.2145*
Exper	.033548*	.0322203*	.0194*	.0220*
Expersq	-.000409*	-.0003958*	-.0002*	-.0002*
Log of Hours p.m.	.1089995*	.1036497*	.1246*	.1250*
Constant	5.543329*	5.601965*	4.838*	4.672*
Lambda	-.139954*	-.25413766*	-.2660*	-.2271*
No of Obs	14 124	11 886	12 723	9 393
Model Chi2	2775*	1687.44*	1939.9*	1284.5*

* Significant at the 1% level ** Significant at the 5% level

The provincial dummies show that African individuals in the Eastern Cape, Northern Cape and Free State in all cases, are likely to earn less than their counterparts in the Western Cape. The differential ranges from about 11% for males in the Eastern Cape to 56% for females in the Free State. The coefficients for both males and females appear to be relatively insensitive to the two unemployment definitions. The Northern Province is the only other province where the results are all significant. However, in this case, the coefficients are all positive. This seems contrary to poverty estimates of the province which place it far below the Western Cape. However, what this may suggest is that for the African employed, the Northern Province offers better earnings potential than the Western Cape. Indeed the mean wage in the Western Cape is only about half that of employees in the Northern Province. The urban-rural estimates add needed detail to this picture. It is not the Northern Province as a whole that offers better earnings but urban employment in the Northern Province. Indeed for rural Northern Province and all other provinces, average male earnings are significantly lower than in the Western Cape.

The sectoral dummies show a strong and clear pattern: relative to Agriculture all the African employed earn more on average. This result holds true for both males and females and according to both definitions. For males, the ranking of the largest wage differentials does not alter by unemployment definition. The sector which pays the most relative to Agriculture is Electricity, where individuals are earning about 88% more than those in farming. This is followed by Transport, Community & Social Services and Finance. The relatively low ranking of Finance, given that it is nationally the highest paying sector, is due to the low representation of African workers here. For females though, Finance does rank higher, although the differential, at about 47%, is lower. The ranking change for females is due to the low ranking of Community services, where females only earn about 36% more than women in Agriculture. This can be explained by the large number of female basic service workers, particularly domestic workers, in this sector. Note that for the two large employers in the economy, Mining and Manufacturing, male workers will tend to earn 60% or more than those in farming, while for females the differential is much smaller at about 25%.

The results by occupation show that for the skilled occupations (Managers, Professionals and Technicians), these individuals are likely to earn between 47% and 76% more than farm labourers. As we move to the semi-skilled occupations (Clerks, Service & sales, Skilled agriculture, craft workers and machine operators) the differentials are smaller. Hence for these occupations individuals earn between 15% and 22% more than farm workers. In the unskilled category though, the results are slightly different and in some cases, surprising. For females, household domestic workers earn about 36% less than farm workers. The coefficient for male labourers in manufacturing though is surprising. Male labourers in manufacturing are seen to earn about 6% less than male farm labourers. For females though, manufacturing labourers earn more. Hence, it would seem that the often perceived higher wage for unskilled workers in the manufacturing industry is driven by the wage differential between women and not men in these two sectors. One can see these same forces and a similar logic operating in the case of domestic helpers. The negative mining labourer coefficient for males (expanded definition) may be reflecting the fact that the mining industry's average skill levels have been increasing in the last decade. Hence those at the bottom have found their wages lagging in preference to those higher up in the internal labour market. Indeed many of the workers in the mining industry would be in the semi-skilled categories.

The union-wage effect is shown here to be about 20% for males and marginally higher at 21% for females. This is substantially lower than the cross-section estimate of Fallon & Lucas (1998) where the differential was over 50%. However their time-series analysis delivered an estimate in the range of 25 to 35%, which is more agreeable with the number here. It cannot be doubted though that union membership is associated with significantly higher earnings for African workers. Tables 7 and 10 reveal that there is a particularly strong union effect in rural areas. The union premium is about 23% for males and 30% for females.

The experience variable indicates that an additional year of experience generates a return to earnings of about 3% for African males. For African females, the return is lower at about 2%. The log of hours worked is significant for both genders and definitions. The coefficients suggest that an increase in the percentage of hours worked will increase earnings by between 10 and 12%. This is quite important as it indicates that an important determinant of earnings is the hours that the African employed are working. Table 7 suggests in urban areas in particular, should males or females opt to work more, the returns could be quite high. This finding is particularly noteworthy as Bhorat and Leibbrandt (1999) show that nearly all of the earners in the

sample are working close to a 40-hour week. Thus, this finding is not contingent on the presence of a significant number of part-time and infrequent workers in the sample.

Finally, as with the employment equation, the Mill's ratio is shown to be significant and negative for all cases. There was therefore a sample selection bias, which was corrected for. The sample of earners are not a random selection of people drawn from the pool of participants. The significance of lambda once again vindicates the selection procedure utilised here.

5. Conclusion

This study has tried to be as meticulous and transparent as possible in modelling the labour market. To this end, the short review of other models highlighted their strengths and drawbacks, while also offering the reasoning for the methodological approach taken here. Perhaps the strongest point to emerge from the methodological section was the insistence on a very carefully managed, three-phase labour market selection procedure, from participation to employment and then to earnings.

The participation equation showed that discouraged workers are statistically closer to the non-participants than to the narrowly unemployed. This strongly suggests that those searching for employment are more likely to get a job than those no longer searching, and therefore hints at the importance of structural unemployment in understanding the participation decision. What makes this so bleak is the fact that many of the youth are in this category rather than in the searching category. Our employment analysis showed that the rural and urban unemployed have different characteristics but similar probabilities of getting employment. What is important about this is that it highlights an asymmetry. Urban work-seekers could take rural jobs but, on average, rural work seekers do not have the characteristics to compete in the urban job market. Rural work-seekers should thus be looking for work in rural areas. This suggests also that the spatial rigidities are essential to understanding employment creation in the domestic economy. The significance of the sample selection terms in the earnings functions also make it clear that those that get employment are different from those that try and do not. The key differences seem to be age and education.

Across the equations then, the age and education variables are important determinants. The age results for the participation and employment equation, in different ways, reflect the importance of youth unemployment. In the participation equation the older age cohorts all have a higher probability of participating than the youth. In turn, the stronger effect in the narrow definition case, points to the significant proportion of youth who are discouraged job-seekers – a fact which has important policy ramifications. The employment probit again suggested that the youth were the least likely to gain employment relative to those in the older age cohorts.

The education results showed very interesting variation across the three equations. Hence, while the non-tertiary education splines tend to be significant and positive in the participation equation, the non-tertiary splines are negative in the employment estimation. This suggests that while non-tertiary education levels tend to increase the probability of participation, these levels are not sufficient to ensure employment. This is a result that matches well with the economy's current, and in all likelihood, future labour demand patterns where firms' specifications are directed primarily toward highly skilled workers in the economy. However it is clear that for those who already have a job, the returns to schooling operate as expected, with secondary schooling yielding a higher rate of return than primary schooling. The fact that we have concentrated so heavily on the vulnerable was shown by the insignificant tertiary coefficient, indicating very low levels of schooling amongst the African workforce. Essentially though, the results across the equations show that education levels operate differentially at each phase of the labour market process.

Ultimately this study has made an additional contribution to the burgeoning earnings function literature for the South African labour market. Through our use of a three-phase model and concentration on the most vulnerable in the labour market, this paper has added value to the literature. In addition the results obtained, particularly in the case of covariates such as location and education, do offer some important background information for policy makers interested in the problems of and solutions to long-term sustainable employment for the domestic economy.

References

Andersson, P. Undated. Labour Market Structure in a Controlled Economy: The Case of Zambia. Ekonomiska Studier Utgivna av Nationalekonomiska Institutionen Handelshogskolan vid Gotemborgs Universitet No 37.

Bhorat,H. 1999. The October Household Survey, Unemployment and the Informal Sector: A Note. *African Journal of Economics*. 67(2): 320-6.

Bhorat, H & J. Hodge, 1999 *Decomposing Shifts in Labour Demand in South Africa, . South African Journal of Economics*. Forthcoming,67(3).

Bhorat,H & M.Leibbrandt, 1998 Poverty amongst the Self-Employed, *Studies in Economics and Econometrics,* 22(3): 25-41.

Bhorat,H & M.Leibbrandt, 1999 Correlates of Vulnerability in the South African Labour Market. *DPRU Working Paper,* No. 99/27: Development Policy Research Unit, University of Cape Town.

Breen, R. 1996. *Regression Models: Censored, Sample-Selected, or Truncated Data*. Sage Publications. Thousand Oaks.

Fallon, P. and R. Lucas. 1998. South African Labour Markets: Adjustment and Inequalities. *Discussion Paper No. 12*, Informal Discussion Papers on Aspects of the Economy of South African. World Bank Southern African Department. Washington.

Fields, G. 1998. *Accounting for Income Inequality and Its Change*. Unpublished Paper. Cornell University. Ithaca.

Glick, P and Sahn, D. 1997. Gender and Education Impacts on Employment and Earnings in West Africa: Evidence from Guinea. *Economic Development and Cultural Change* 22: 95-125.

Greene, W.H. 1993. *Econometric Analysis,* 2nd Edition. Macmillan Publishing Company. New York.

Heckman, J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47: 153-161.

Heckman, J and Hotz, J. 1986. The Sources of Inequality for Males in Panama's Labor Market. *The Journal of Human Resources* 21(4):507-542.

Hofmeyr, J. 1998. Segmentation in the South African Labour Market. *Working Paper No. 15*. South African Network for Economic Research. Potchefstroom.

Hofmeyr, J and Lucas, R. 1998. The Rise in Union Wage Premia in South Africa. *IED Discussion Paper Series No. 83*. Institute for Economic Development, Boston University.

International Labour Office. 1996. *Restructuring the Labour Market: The South African Challenge*. ILO Publication Office. Geneva.

Lipton, M., de Klerk, M. and Lipton, M. 1996. *Land, Labour and Livelihoods in Rural South Africa*. Volume One. Indicator Press. Durban.

Moll, P. 1998. Discrimination is Declining in South Africa but Inequality is not. *Working Paper No. 5*. South African Network for Economic Research. Potchefstroom.

Mwabu, G and Schultz, T. 1996a. *Wage Premia for Education and Location by Gender and Race in South Africa*. Unpublished Paper. Yale University. New Haven.

Mwabu, G and Schultz, T. 1996b. "Education Returns Across Quantiles of the Wage Function; Alternative Explanations for the Returns to Education by Race in South Africa". *American Economic Review* 86(2): 335-339.

Mwabu, G and Schultz, T. 1997. *Labour Unions and the Distribution of Wages and Employment in South Africa*. Unpublished Paper, Yale University. New Haven.

Nattrass, N. and Seekings, J. 1998. Changing Patterns of Inequality in the South African Labour Market. In, Petersson, L. (ed.). *Post Apartheid Southern Africa*. Routledge. London.

Roy, A. 1951. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*. 3: 135-146.

Winter, C. 1998. *Women Workers in South Africa: Participation, Pay and Prejudice in the Formal Labour Market*. Unpublished Paper. World Bank.

Appendix

Table 5 : Urban African Male and Female Labour Participation Equations for Expanded and Narrow Definitions of Unemployment

	Urban Male				Urban Female			
	Expanded		Narrow		Expanded		Narrow	
	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar
None-Std5	.00416*	5.057	.00770*	5.05754	.008409*	5.05748	.00715**	5.057
Std 6-10	.00675*	1.857	.01722*	1.85798	.03802*	1.8025	.06277*	1.8026
Tertiary	-.0012	.1366	.00080	.136614	-.012424	.14637	.012587**	.1463
26-35	.04769*	.3700	.1295*	.370018	.11867*	.37385	.17698*	.3738
36-45	.05944*	.2720	.1759*	.272077	.12216*	.26495	.24484*	.2649
46-55	.04078*	.1420	.14104*	.142091	.071356	.14082	.22022*	.1408
56-65	.02572*	.0484	.12731*	.04884	.01957	.04308	.19905*	.0430
No. of Kids <7	.00653	.6804	.0138*	.680472	-.01264*	.93524	-.02184*	.9352
No. of Kids 8-15	.00137	.7273	-.0025	.727369	.00071	.94158	-.00553	.9415
No of males 16-59	-.00198	1.989	-.0231*	1.98956	-.00872	1.3584	-.01304*	1.358
No of fems 16-59	-.01114*	1.461	-.0256*	1.46165	.0211*	2.12692	.01154**	2.126
No of Adults >60	-.02709*	.2766	-.06931*	.276624	.00075	.28785	-.01897*	.2878
Other hhld income	-4.05e-07*	22205.2	-3.63e-07	22205.2	-2.12e-06*	26645.	-1.96e-06*	26645
Other hhld income sqred	7.98e-13	1.5e+09	1.80e-12	1.5e+09	2.33e-12*	2.8e+09	2.15e-12*	2.8e+09
Obs. Prob	**.9367**		**.8243**		**.7925**		**.6377**	
Pred. Prob (at x-bar)	**.9519 (at x-bar)**		**.8540 (at x-bar)**		**.8105 (at x-bar)**		**.6508 (at x-bar)**	
No Obs	**6521**		**6521**		**7707**		**7707**	
Chi(2)	**328.2***		**908.61***		**548.9***		**785.33***	
Pseudo R2	**0.098**		**0.1386**		**0.0665**		**0.076**	

* Significant at the 1% level ** Significant at the 5% level

Table 6: Urban African Male and Female Employment Equations for Expanded and Narrow Definitions of Unemployment

	Urban Male				Urban Female			
	Expanded		Narrow		Expanded		Narrow	
	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar
None-Std5	-.01593*	5.085	-.01438*	5.108	-.00227	5.1671		
Std 6-10	.01392**	1.883	.00231	1.914	.06397*	1.9462		
Tertiary	.03879*	.1395	.042044*	.1470	.11611*	.16291		
26-35	.05078*	.3749	-.02919	.3745	.28184*	.39632		
36-45	.12273*	.2817	-.01835	.2987	.42543*	.27488		
46-55	.16620*	.1437	.03033	.1512	.39832*	.13170		
56-65	.23943*	.0461	.08335*	.0499	.39031*	.03621		
E.Cape	-.02294	.0938	.06925*	.0862	-.06509**	.13070		
N.Cape	-.05055	.0124	.00093	.0119	-.10312**	.01186		
Free Stat	-.02629	.0841	.10175*	.0768	-.07232**	.10782		
Kwaz/Natl	.01830	.1522	.02675	.1534	-.00886	.1727		
North-W	.07827**	.0809	.06247*	.0828	.02066	.06899		
Gauteng	.0513**	.4459	.06451*	46052	-.00559	.38568		
Mpumal	.03884	.0391	.07693*	.0362	-.15908*	.0366		
N.Prov	.04587	.0258	.1129*	.0242	-.11747*	.02743		
Lamda	-1.3549*	.1206	-.56492*	.2726	.18644*	.33571		
Obs. Prob	.7329		.8328		.6061			
Pred. Prob At x-bar)	.7576 (at x-bar)		.8610 (at x-bar)		.6248 (at x-bar)			
No Obs	6056		5206		5957			
Chi(2)	1082.7*		681.4*		1148.7*			
Pseudo R2	0.1481		0.1407		0.1427			

* Significant at the 1% level `** Significant at the 5% level

Table 7: Urban African Male and Female Earnings Equations for Expanded and Narrow Definitions of Unemployment

	Urban Male		Urban Female	
	Expanded Unemployment	Narrow Unemployment	Expanded Unemployment	Narrow Unemployment
None-Std5	.04804*	.0499*	.0426*	.0448*
Std 6-10	.1071*	.1052*	.0741*	.0924*
Tertiary	.05094	.0447	.0343	.0485
E.Cape	-.03586	-.0455	-.1482*	-.1700*
N.Cape	-.0574	-.0407	-.2584*	-.2725*
Free Stat	-.21795*	-.2336*	-.3370*	-.3707*
Kwaz/Natl	.0883	.0893**	-.0624	-.0609
North-W	.1077	.1017	-.1258*	-.129**
Gauteng	.1599*	.1581*	.1697*	.1629*
Mpumal	.08287	.079	.1065	.0643
N.Prov	.32880*	.3037*	.0759	.0555
Mining	.1661762*	-.1603029*	.4207	.4217
Manuf	.2845647*	.2828365*	.2309	.2262
Electricity	.5899361*	.5866131*	.533	.5350
Constr	.1137356	.1129959	.3380	.3297
Wholes	.1247791	.1256877	.1888	.1848
Transport	.372437*	.3683933*	.4944*	.4952**
Finance	.3085486*	.3042547*	.4715**	.4702**
Comm Serv	.3275005*	.3256818*	.3941**	.3899
Other	.227513	.020014	.1066	.1224
Armed Forces	.8778821*	.8646148*	.7430	.7425
Managers	1.112009*	1.114832*	.929*	.9507*
Profess	0.9747265*	.9663315*	.9155*	.9360*
Technicians	.7456466*	.7399523*	.7825*	.7989*
Clerks	.5202852*	.5163703*	.5080*	.5249*
Serv&Sales	.4786886*	.4743847*	.2533	.2686
Skilld Agric	-.073708	-.0684681	.6278	.6321
Craft	.4804153*	.4768006*	.0935	.1083
Mach Operator	.470309*	.4668909*	.2986**	.3162**
Unspecif	.3251987*	.3239393*	.4489*	.4670*
Domes Helper	.2419549	.2379359	.0871	.1027
Mining lab	.2490426	.2472353	-.0440	-.0381
Manuf lab	.3638321*	.3604891*	.2534	.2663
Trprt lab	.2505165	.2492946	-.1385	-.1355
Domes Worker	-.4324948*	-.4373609*	-.4373**	-.4223
Union Member	.163868*	.1641*	.1741*	.1768*
Exper	.02947*	.0282*	.0122*	.0174*
Expersq	-.00032*	-.0003*	-.0001*	-.0001*
Log of Hours p.m.	.1803*	.1736*	.1623*	.1629*
Constant	4.8883*	4.969*	5.065*	4.809*
Lambda	-.1288*	-.2411*	-.2774*	-.1604
No of Obs	6018	5185	5922	4689
Model Chi2	1142.33*	755.44*	1185.7*	737.16

* Significant at the 1% level ** Significant at the 5% level

Table 8 : Rural African Male and Female Labour Participation Equations for Expanded and Narrow Definitions of Unemployment

	Rural Male				Rural Female			
	Expanded		Narrow		Expanded		Narrow	
	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar
None-Std5	.00339**	3.794	.00461	3.794	-.00147	3.664	.00169	3.664
Std 6-10	.00298	1.017	.0146*	1.017	.05909*	.9386	.05025*	.9386
Tertiary	.00721	.0745	.03130**	.0745	-.03539**	.0762	.00267**	.0762
26-35	.06557*	.3282	.15007*	.3282	.07492*	.3224	.09867*	.3224
36-45	.08092*	.2248	.19841*	.2248	.03958**	.2442	.13210**	.2442
46-55	.05995*	.1442	.17883*	.1442	-.07081*	.1543	.07275*	.154
56-65	.00657	.0580	.15387*	.0580	-.21457*	.0547	-.01019*	.0547
No. of Kids <7	-.00131	.8929	.0026	.8929	-.02523*	1.228	-.02566*	1.228
No. of Kids 8-15	-.0022	1.032	-.01241*	1.032	-.01393*	1.271	-.01911*	1.271
No of males 16-59	-.01229*	1.938	-.0295*	1.938	-.01024	1.323	-.01518	1.323
No of fems 16-59	-.00456**	1.568	.01966*	1.568	.03333*	2.166	.01238*	2.166
No of Adults >60	-.03176*	.3625	-.10120*	.3625	-.00143	.3831	-.03698	.3831
Other hhld income	-2.19e-06*	12414	-5.10e-06*	12414	-1.94e-06*	17180	-4.14e-07	17180
Other hhld income sqred	1.96e-11*	7.1e+08	5.18e-11*	7.1e+08	1.98e-12*	1.6e+09	8.38e-14	1.6e+09
Obs. Prob	**.8862**		**.7255**		**.5723927**		**.3895**	
Pred. Prob (at x-bar)	**.9230 (at x-bar)**		**.7774 (at x-bar)**		**.5775 (at x-bar)**		**.3849 (at x-bar)**	
No Obs	**9137**		**9137**		**11841**		**11841**	
Chi(2)	**842.47**		**1777.8***		**1009.3***		**832.2**	
Pseudo R2	**0.1344**		**0.1686**		**0.0626**		**0.0523**	

* Significant at the 1% level ** Significant at the 5% level

Table 9 : Rural African Male and Female Employment Equations for Expanded and Narrow Definitions of Unemployment

	Rural Male				Rural Female			
	Expanded		Narrow		Expanded		Narrow	
	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar	Marginal Effects	x-bar
None-Std5	-.00724*	3.8013	-.00365	3.7785	.0025	3.8942	-.00586	3.9199
Std 6-10	.0059	1.0072	-.00543	1.0055	.00363	1.1610	-.0602*	1.2250
Tertiary	.0519*	.0751	.0341**	.08079	.15577*	.09371	.18805*	.11438
26-35	.02973	.33754	.01835	.33751	.17383*	.36172	.05610*	.35061
36-45	.06921*	.24030	.03011	.25760	.31589*	.24621	.10424*	.26615
46-55	.12879*	.14988	.04884*	.16165	.40319*	.12676	.21521*	.14366
56-65	.26790*	.05208	.09339*	.05849	.44982*	.03099	.24965*	.03773
E.Cape	-.36825*	.15302	-.09449	.1333	-.34332*	.19162	.06836	.1821
N.Cape	-.00751	.0087	-.12155	.01068	.11547	.00297	.10189	.00417
Free Stat	-.12948	.09123	-.00748	.10507	-.07559	.07263	-.02837	.08728
Kwaz/Natl	-.20595**	.23660	-.08260	.23725	-.26387*	.26490	.0877	.27169
North-W	-.31503*	.15361	-.07203	.15270	-.28464*	.12382	.09488	.11817
Gauteng	.0149	.03873	-.00889	.04601	-.00525	.0234	-.03883	.03088
Mpumal	-.27706*	.15002	-.02809	.15217	-.28012*	.1170	.06296	.11361
N.Prov	-.17699	.15103	-.01851	.14237	-.34838*	.19861	.10333	.18577
Lamda	-1.1393*	.18971	-.31876*	.38249	-.14034	.64426	-.65184*	.93297
Obs. Prob	.7005		.8556		.4926		.7239	
Pred. Prob At x-bar)	.7357 (at x-bar)		.8946 (at x-bar)		.4953 (at x-bar)		.7524 (at x-bar)	
No Obs	8147		6725		6853		4721	
Chi(2)	1838.6*		986.15*		978.65		678.4*	
Pseudo R2	.1882		.1840		.1030		.1236	

* Significant at the 1% level ** Significant at the 5% level

Table 10: Rural African Male and Female Earnings Equations for Expanded and Narrow Definitions of Unemployment

	Rural Male		Rural Female	
	Expanded Unemployment	Narrow Unemployment	Expanded Unemployment	Narrow Unemployment
None-Std5	.0262*	.0273*	.0478*	.0488*
Std 6-10	.1004*	.1013*	.0890*	.0922*
Tertiary	.0114	.0102	.0371	.0048
E.Cape	-.4011*	-.4103*	-.0656	-.0444
N.Cape	-.4151*	-.4085*	-.1938	-.1476
Free Stat	-.5753*	-.5771*	-.7315*	-.715*
Kwaz/Natl	-.2224*	-.2248*	.1796	.2003
North-W	-.3300*	-.3336*	-.0327	-.0311
Gauteng	-.3541*	-.3568*	-.0183	.0089
Mpumal	-.3225*	-.3338*	.1668	.1804
N.Prov	-.2000*	-.2089*	.2416	.2591**
Mining	.7579127*	.7570094*	.2203	.2244
Manuf	.6252782*	.6272172*	.2482**	.2595*
Electricity	.8139617*	.8148491*	.5340	.5443**
Constr	.5345509*	.536156*	.4242**	.4268*
Wholes	.5716519*	.5738369*	.1798	.1843**
Transport	.7709761*	.7735797*	.5935*	.5932*
Finance	.5965777*	.5997498*	.4253*	.4230*
Comm Serv	.7018911*	.7037523*	.3021*	.3087*
Other	-.3265729*	-.32589*	.3742*	.3783*
Armed Forces	.49524*	.493589*	(dropped)	(dropped)
Managers	.6469359*	.6474902*	.8197*	.8297*
Profess	.9065049*	.9038113*	1.104*	1.097*
Technicians	.5282483*	.5276259*	1.022*	1.020*
Clerks	.2504032*	.2509869*	.6078*	.6127*
Serv&Sales	.141053*	.1413896*	.3725*	.3735*
Skilld Agric	.3405803*	.3402449*	-.0404	-.0176
Craft	.1922901*	.1922264*	.2898*	.2979*
Mach Operator	.1111536*	.1112566*	.2742**	.2817**
Unspecif	-.0562722	-.0549814	-.0151	-.0055
Domes Helper	-.0486589	-.0478983	.2741*	.2729*
Mining lab	-.1081574**	-.106682**	.2514	.2532
Manuf lab	-.2533656*	-.2527708*	.1950	.1891
Trprt lab	-.0079984	-.0101031	-.6912	-.6433
Domes Worker	-.8721786*	-.8712416*	-.2885*	-.2936*
Union Member	.2366*	.2359*	.2969*	.2947*
Exper	.0397*	.0392*	.0272*	.0243*
Expersq	-.0005*	-.0005*	-.0003*	-.0003*
Log of Hours p.m.	.0324	.0307	.1035*	.1012*
Constant	6.172*	6.179*	4.695*	4.752*
Lambda	-.0889**	-.1321*	-.1706*	-.3425*
No of Obs	8106	6701	6801	4704
Model Chi2	1927.80*	995.55*	1026.9*	708.2*

* Significant at the 1% level ** Significant at the 5% level