



**MONASH** University

**Australia**

Department of Econometrics  
and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Language and Labour in South Africa  
A new approach for a new South Africa**

**Katy Cornwell**

**February 2006**

**Working Paper 5/06**

# Language and Labour in South Africa

*A new approach for a new South Africa*

January 2006

**Katy Cornwell**

Department of Econometrics and Business Statistics

Monash University

Clayton VIC 3800

Australia

Phone +61 3 9905 2453

Fax +61 3 9905 5474

[katy.cornwell@buseco.monash.edu](mailto:katy.cornwell@buseco.monash.edu)

## *Abstract*

---

This paper considers the role of language in labour earnings in South Africa over the period 1996 to 1998. Our pooled cross-section comprises of over 160,000 working age adults, and the analysis considers the decision to participate in the labour force, employment prospects and labour earnings. Models include variables for individual mother tongue in addition to population group. After conditioning on a number of socio-economic and demographic factors, we find that having English as one's mother tongue is one of the pivotal determinants of labour earnings. These results are robust across two models of sample selection. Such findings shed light on the economic consequences of South Africa's national policy of linguistic heterogeneity.

Keywords: Unemployment, Income, South Africa, Language Policy, Race

**JEL Codes: J64, J21, J23, J19, J31**

---

The author would like to thank Brett Inder and Ana Deumert for useful discussions.

## 1. Introduction

Unemployment in South Africa has been metaphorically described as an untamed beast (Kingdon and Knight, 2004). Indeed, South Africa is characterised by unemployment rates amongst the highest in the world, with the most ferocious rates of up to 45% amongst black South Africans (see Table 1). Moreover, hardship is not overcome once paid employment is found, particularly for the majority black South African population. As a residue from the apartheid era, substantial differences in earnings between racial groups remain. Table 1 demonstrates that on average white South Africans are earning almost four times as much as blacks.

High unemployment rates coupled with disparate labour earnings have led to numerous studies on the determinants of income, mostly focussing on the returns to education. The latest studies include the work of Keswell and Poswell (2002) and Serumaga-Zake and Naude (2003). The former work questions the empirical relevance of the standard human capital theory of diminishing marginal returns to education, and provides a thorough overview of the vast South African returns to education literature. Serumaga-Zake and Naude (2003) utilise double hurdle and Heckman sample selection models in examining the private returns to education of black South African males and females. We extend this work by considering multiple years of data from the South African October Household Surveys and incorporating additional variables.

Apartheid dictated that population group was the primary determinant of educational and occupational opportunity. With well-documented evidence of the effect of education on earnings, it was natural for the literature to link population group, education and earnings in South Africa. The breakdown of apartheid saw the formation of the new South Africa where black economic empowerment is recognised as “fundamental to redressing past imbalances and enabling the country to move on to achieve sustainable development and prosperity” (southafricainfo, 2004). Great efforts have been made to eradicate racial discrimination and undo the injustices of the past, and the Rainbow Nation looks forward to the day when they can say with confidence that population group no longer determines one’s fate. Disappointingly, studies continue to find population group dummies strongly significant in income and employment equations. We would argue that a new South Africa calls for a new approach to modelling the South African labour market: an approach which looks much further than population group in identifying the determinants of earnings in the multilingual new South Africa.

We begin to explore this notion by examining whether mother tongue language provides a better insight than population group into what is of importance to an individual’s relative success in the labour market. Is it population group *per se* that leads to higher unemployment rates for black South Africans, or is it that

English is not their natural mother tongue language, creating a barrier of entry to employment and an impediment to earnings? It is this aspect of the labour market which we seek to address in this paper.

Our interest is primarily on introducing language as a potential determinant of labour earnings. However, before the individual is able to report earnings, they must overcome two hurdles: the individual must first choose to participate in the labour force, and then from this labour force pool the individual must also be selected for employment. Recognising a propensity for sample selection bias, we model income using two models of sample selection: a version of Cragg's (1971) double hurdle and Heckman's (1979) sample selection model, with mother tongue included in addition to population group and other socio-economic and demographic variables at each of the participation, employment and income stages of the models.

Given that this avenue is a new direction for the labour earnings literature, our next section is devoted to discussion of the South African labour market in the context of language. Section 3 follows with a description of the methodology, while section 4 introduces the data. Results are presented in section 5, and discussion follows in section 6.

## **2. Language as the new direction**

Embracing linguistic pluralism in its constitution, the new South Africa recognises and guarantees equal status to each of its eleven official languages<sup>1</sup>. However, historical white dominance in government and commerce is reflected in English and Afrikaans being the most commonly used languages in official and commercial public life, despite the African languages of Xhosa and Zulu being the more common languages spoken at home (see Table 2). In particular, English is eighth on the list of mother tongues ranked according to frequency for the respondents in our sample.

Literature on the economics of language in the labour market is limited. The majority involve consideration of the role of language in labour market interactions and earnings for immigrants and Hispanics in the United States. Discussion tends to find consensus in favour of linguistic homogeneity.

The theoretical basis for the immigrant/Hispanic literature is generally pinned to the notion of language as the facilitator of communication. In this sense, language can be seen as the medium for communication exchange, whereby linguistic heterogeneity increases the transaction costs of this exchange and consequently in the absence of bilingualism, less exchange will take place between those speaking different languages.

Consider the implications of this for the individual job seeker in the labour market. Information about jobs flows through open channels of communication. The individual is privy to these channels of information depending on her ability to communicate with the people in these channels. If English is the dominant language used in the work environment, the English-speaking individual can tap into information directly from the pool of the employed and also directly to the employer. Consequently, the English-speaker holds an advantage on the employment front over the non-English speaker. This suggests that job search may not only be facilitated by language channel, but also limited by it, and hence linguistic disadvantage would present itself in both the participation decision and the employment outcome.

The literature advocates that earnings may also be tied to language knowledge. Where access to occupation is determined by language channel, a worker may find themselves in low paid occupations relative to skill. Kossoudji (1988) suggests that there could also be some element of individual choice to be among peers of the same language background, thereby maximising individual utility rather than income. Indeed, this could be one explanation for why large income differences are observed by language group (see Table 2).

From the employer's viewpoint, communicative ability is a form of human capital in that it enhances productivity, and productivity is linked to earnings. The communicator is better able to convey their comparative skill advantage, from which the employer is able to realise any productivity gains from specialisation. McManus (1985) compares a group of employees sharing a lingua franca with a group who are unable to communicate. He suggests that in the latter case, the division of tasks would be according to average characteristics of the group with no allowance for personal variation, whereas the former would be more productive in the sense that productivity gains from specialisation could be realised. Hence, those with high levels of communication would find themselves deemed more productive and awarded accordingly with higher income.

Furthermore, by accelerating the absorption of information, communication improves the return to education (McManus *et al* 1983), such that employers may choose to train workers with high language skills in new technology more readily than those with limited language skills, enabling the employee with high language skills to climb further up the promotional and therefore income ladder.

Empirical application corroborates this theoretical discussion, revealing that language attributes play an important role in earnings for immigrants and Hispanics. Grenier (1984) is able to use language to explain up to one third of the relative wage difference between Anglo and Hispanic men. Kossoudji (1988) concurs in her selection bias corrected specification of a random utility model for occupation and earnings. Other studies have been able to incorporate English proficiency. For instance, Rivera-Batiz (1990) uses test-based

---

<sup>1</sup> The 11 Official languages of the Republic of South Africa are Sepedi, Sesotho, Setswana, siSwati, Tshivenda, Xitsonga, Afrikaans, English, isiNdebele, isiXhosa and isiZulu.

English proficiency measures to examine the impact on earnings, finding it to be a major factor. Mora (2003) models a standard Mincerian earnings function with English fluency, geographical region and ethnicity as conditioning variables, finding there to be a great deal of interaction between education, experience and schooling. Interestingly, for males with no education and no experience, Mora (2003) suggests that those who speak English earn significantly less than those who do not speak English, yet this result is reversed at higher levels of schooling: for a male with 12 years of education, English language proficiency adds an earnings premium of 30%.

Small pockets of studies have also looked at the effect of minority languages on educational outcomes in developing economies, however, while interesting, these are mainly descriptive and their emphasis on the implications for bilingual education distracts from the focus of our study concerning labour market outcomes. Moreover, like the work on immigrants, these studies are concerned with the implications of a minority population group being unable to converse in one official and dominant language. The case of South Africa, however, is unique in that through a history of political discrimination, the languages of the minority population group dominate commerce and official life, yet it is the majority – and also the poorest – population group who speak the languages which could well be the “minor” in the labour market context. We turn now to our own analysis in the hope of shedding light on this situation.

### **3. Methodology**

Our analysis is primarily concerned with the factors which contribute to monthly earnings for all South Africans. Prior to reporting income, the individual must first choose to participate in the labour force. Of course, in South Africa participation in the labour force does not guarantee employment, and so a further decision on behalf of the employer must be made to draw the individual from the labour force pool. Only once the individual is employed do they report income. Accordingly, only a subsample of all South Africans are employed and able to report earnings. It is likely that the socio-economic characteristics of the employed are different to those who are not, and likewise, the characteristics of labour force participants are different from non-participants. In particular, unobservable characteristics affecting the decision to work would be correlated with the unobservable characteristics affecting income. Selectivity bias would arise, therefore, if we were to make statements about the determinants of earnings for all South Africans based on the observed earnings of the subset whom are employed. The appropriate model must be one which copes with sample selection at each stage of participation and employment.

We specify a sample selection model with the primary dependent variable of interest of the form

$$y_i^I = x_i^I \beta^I + u_i^I,$$

where  $x_i^I$  is the vector of socio-economic and demographic explanatory variables,  $\beta^I$  the vector of unknown coefficients and  $u_i^I$  the error term.

There are two latent decision functions:

- (1) The participation decision:

$$I_i^{P*} = x_i^P \beta^P + u_i^P,$$

with indicator variable

$$I_i^P = \begin{cases} 1 & \text{if } I_i^{P*} > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Such that the individual participates in the labour force if  $I_i^{P*} > 0$ .

And

- (2) The employment decision:

$$I_i^{E*} = x_i^E \beta^E + u_i^E,$$

such that the individual is selected for employment if  $I_i^{E*} > 0$  and hence

$$I_i^E = \begin{cases} 1 & \text{if } I_i^{E*} > 0 \\ 0 & \text{otherwise} \end{cases}.$$

These choices are partially observed: we do not observe the employment outcome for the non-participant nor the income for unemployed and non-participants. Correspondingly, the participation equation is defined over the whole South African working age population, the employment equation over labour force participants and the income equation over those who are employed.

We utilise 2 models of income determination which incorporate sample selection correction factors: the double hurdle model and Heckman's sample selection model with two sample selection mechanisms.

### ***Double hurdle model***

The double hurdle literature is divided about whether the participation and employment decisions are joint or sequential – that is, does one's perceptions of their employment prospects influence their decision to participate in the labour force, or is the participation decision taken first, independently of employment?

Maddala (1993) embraces the joint versus sequential debate by pointing out that the distinction rests in the covariance of error terms in the participation and employment equations. As extensions to the Heckman-Lee two stage estimation methods, Maddala (1993) outlines a sequential double hurdle model with uncorrelated errors, and then relaxes the zero correlation assumption to describe a joint double hurdle model relying on bivariate probit estimation. Smith (2002) renders joint double hurdle approaches superfluous, arguing that the specification of Cragg's original model is already a joint model. Confusion as to whether decisions are joint or sequential, he purports, stem from the ease with which Cragg's model lends itself to sequential interpretation. Such unresolved ambiguity ultimately leaves the choice between joint and sequential models in the hands of the researcher.

To add a further complexity to the issue, South Africa has long debated on the particular definition of the unemployed: under the official (narrow) definition, working-age individuals who are not employed must be actively searching for work and be able to start work within 1 week in order to be deemed unemployed. An alternative definition (the so-called broad or expanded definition) relaxes the need for the individual to be actively searching, and to some extent allows inclusion of discouraged workers. The difference that this criterion makes for the unemployment rate is quite substantial (see Table 1). For comparability and completeness, our analysis is undertaken using both definitions. The issue of correlation between employment and participation is somewhat intertwined with these definitions of unemployment. The difference in results owing to choice between Cragg and Maddala, or joint and sequential is likely to be minimal, hence we opt for the computational simplicity of Maddala's (1993) sequential model.



Assuming normality of the error terms, Maddala's (1993) double hurdle model involves estimating two separate probit models for participation and employment. From these estimated models we obtain two correction factors

$$\lambda_i^P = \frac{\phi(x_i^P \hat{\beta}^P)}{\Phi(x_i^P \hat{\beta}^P)}$$

and

$$\lambda_i^E = \frac{\phi(x_i^E \hat{\beta}^E)}{\Phi(x_i^E \hat{\beta}^E)},$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  are, respectively, the probability density and cumulative distribution functions of the standard normal distribution.

Restricting the sample to those employed, income is regressed on a number of socio-economic and demographic variables (outlined in section 4) as well as both the obtained correction factors.

### ***Heckman's sample selection model***

Heckman's model differs from Maddala's (1993) sequential double hurdle model in its inclusion of the two correction factors in estimation. The Heckman participation probit and its corresponding correction term are identical to those of the double hurdle. In modelling the employment probit, however, the correction factor from the participation equation is included as an additional variable for the Heckman model. The second correction factor is then obtained as

$$\lambda_i^E = \frac{\phi(x_i^E \hat{\beta}^E)}{\Phi(x_i^E \hat{\beta}^E)},$$

where  $x_i^E$  now includes  $\lambda_i^P$  as an additional variable. The second correction factor alone is then included in the income equation as an additional regressor.

#### 4. Data

Data for this study is extracted from the South African October Household Surveys of 1996 through to 1998. This provides a multi-stage cluster sample of some 166,534 working age adults.

Our double hurdle and Heckman models characterise income as the primary continuous dependent variable of interest. Participation in the labour market and employment are two hurdles which must be overcome before an individual is observed as recording income. Adopting a Mincerian form to our dependent variable, income is taken as the natural logarithm of real monthly income and deduced from waged and/or self-employed sources.

We define “success” in the labour market as one having a full-time job and earning a consistent wage, and that anyone working less than full-time throughout the month is doing so due to lack of work rather than by choice. Given this definition, we do not adjust monthly income according to hours worked, since total income earned during the month would capture relative “success”.

Monthly income is deflated according to temporal, provincial and regional (rural/urban) price differences. Such a price deflator does not exist for South Africa, and neither can it be constructed from the official Stats SA CPI publications: CPI series are available at the provincial level, but for each province the series is constructed to have a base of 100 in 1998, rendering the series incomparable across provinces. Moreover, the prices used to construct these indices were collected at only the urban level, and do not provide any information on prices in the rural areas. We thus construct our own price deflator by using cluster (village) level price data in the World Bank’s 1993 LSMS dataset in addition to the official provincial urban series.<sup>2</sup>

The first stage of constructing the price indices involves using 1996 official Stats SA CPI weights to calculate a food price index for 1993 by (new) province and region, relative to a base of 100, representing the urban overall country average. The official urban provincial level CPI’s were then rebased to their 1993 food price index value, and re-projected out to 2005, creating an urban provincial level series relating prices back to the 1993 urban country average. The rural counterpart was then constructed at the provincial level according to the within province rural-urban food price difference observed in the 1993 LSMS data. This construction provides a deflator series for the whole country, mapping time, province and region relative to one base: average 1993 urban food price.<sup>3</sup> Its reliability will depend on the rural-urban price differences that prevailed in 1993 remaining stable over time. This may not be the case, but given the lack of alternative data, this represents the best available approximation – it is certainly better than ignoring differences in prices between rural and urban areas.

---

<sup>2</sup> A similar method was used by Kingdon and Knight (1999) to construct spatial price indices for South Africa in 1993.

Participation and employment are the secondary binary dependent variables of interest. In the participation equation, the dependent variable takes a value of 1 where the individual participates in the labour force and 0 otherwise. Similarly, the dependent variable in the employment equation takes a value of 1 where the person is employed. For this second binary variable, the sample is restricted to labour force participants.

The October Household Survey for 1996 introduced a new question regarding the mother tongue of the respondent. Given that South Africa embraces eleven official languages, it is of particular interest whether mother tongue language influences labour market outcomes, particularly employment and income which are determined by the employer. We also condition on a number of socio-economic and demographic variables including population group, gender, household head, rural residency, age and time dummies. Linear regression splines for highest level of education attained are used to allow for differing slopes across primary, secondary, and tertiary education. For the income equation, we also include dummy variables for employment in the informal sector and self employment. The marital status dummy is interacted with individual gender in the participation equation, but omitted from the employment and income equations for identification. A more comprehensive description of the variables is provided in Table 3.

In each of the double hurdle and Heckman models, we consider unemployment under both the official and expanded definitions, with little variation in results.

## **5. Results**

### **5.1 Labour force participation**

Table 4 presents results for the participation decision under both the official and broad definitions of unemployment. Defined over the entire working-age population, the estimated double hurdle and Heckman models are identical in this first stage of participation.

The models in columns (1) and (2) follow the convention in the literature for studies on the South African labour market: inclusion of dichotomous indicator variables for apartheid-era racial classification of the individual, alongside other socio-economic characteristics such as status in the household, gender, age, education etc. As Table 4 reveals, our results corroborate the common findings of such literature, with all population group dummies strongly significant. Under the narrow definition of unemployed (column (1)), we find that Asian, coloured and white South Africans are all more likely to participate in the labour market than are black South Africans.

---

<sup>3</sup> The price index data is available from the author on request.

Males and those regarded as head of the household are more likely to enter the labour force than females and those not regarded as head of their household respectively. Where the male is also married, there is an additional likelihood that they will participate, yet this is reversed for married females. Moreover, age has a quadratic effect on labour force participation, falling as one approaches old age. These results highlight a strong patriarchal culture for South Africa and the importance of the household as an economic and cultural unit: the decision to participate in the labour force is made in conjunction with the cultural obligations of the (often extended) family.

Educational attainment is important for labour force participation. Under the narrow definition, attainment of primary school level education has no effect on entry into the labour force over no education. As one progresses through secondary school and tertiary studies, entry into the labour force becomes increasingly more common. The slope of the linear spline eases off for those with other levels of education, yet remains steeper than that for secondary education. Indeed, higher schooling may improve access to information about employment, enhance one's perception of their employability and lead to a higher rate of participation in the labour force.

Evidence suggests a higher labour force participation rate amongst urban dwellers, reflecting the lack of option for subsistence farming that the rural area provides. Finally, there is some evidence of labour force expansion for South Africa as a whole over the three years, with some provincial variations.

The move from the official to broad definition means that around 18,000 individuals move from being non-participants to unemployed participants. These people could perhaps be called "disheartened workers" as opposed to discouraged workers: they would work given the opportunity, but are not actively seeking work. The results, in column (2), do change under this alternative definition. In most cases the change is only in significance, but for two of the population group and one of the provincial dummies, the estimated coefficients change sign. For those officially classified as Asian or white, a significant positive estimate under the narrow definition becomes a significant negative under the broad definition. Indeed, 92% of the so-called disheartened workers are black South Africans, hence the reclassification boosts the participation rate amongst black South Africans, *ceteris paribus*. Moreover, the now negative coefficient on the white dummy reflects the "old money" syndrome: typically white South Africans have historical wealth for use as a safety net, enabling them to drop out of the labour force much more easily than the invariably poorer black South African.

Interestingly, the coefficient on primary schooling becomes stronger under the broad definition such that those with primary school attainment are less likely to participate in the labour market than those with no schooling. This could imply that a greater proportion of the disheartened have attained only primary school

level education. Alternatively, with the model defined over the entire working age population (15-65), and school attainment not necessarily implying the individual has ceased school, we may be inadvertently capturing those currently in school and hence not participating in the labour force at this stage.

Columns (3) and (4) present results for our alternative variable specification: inclusion of mother tongue language. We group the official African languages together to form the base language category, and use dummy variables to distinguish English, Afrikaans and other (unofficial) languages. We also interact English and Afrikaans with population group to allow for a differing effect of the commerce-dominant languages according to population group.

The results are practically identical for all variables except the population group dummies. In particular, we see a drastic change in the importance of population group for labour force participation once individual mother tongue language is taken into account. These results imply a new mindset for a new South Africa: racial discrimination no longer dominates, rather, the ability of the individual to tap into the world of the employed.

Looking more closely at the results for language, it seems that while coloured South Africans generally have high labour force participation rates which remain unchanged by language, those who have English as their mother tongue are much less likely to participate. Differential participation rates according to language are seen most strongly for the base population group: black South Africans. In particular, under the narrow definition of unemployment displayed in column (3), having English as a mother tongue improves the chances of the average black South African being economically active by around 14%.

For completeness, we also present results based on black South African working age adults alone in Table 5. By concentrating on only the black population, we are able to further disaggregate the official African languages. We construct a new set of 7 dummy variables for language based on prevalence: 6 dummies representing the 6 most common languages to be spoken in the home, with the remaining languages grouped together in the 7<sup>th</sup> dummy as “minority languages”. This time, we use English as the base in order to determine just how much difference there is between the widely spoken languages and that which dominates commerce.

Under the narrow definition, all the African and minority languages are associated with lower labour force participation rates. An Afrikaans mother tongue, however, is not statistically different to an English mother tongue for participation rates amongst black South Africans. In contrast, the broad definition provides much weaker results, with no significant difference in participation according to language. Indeed, the link between language and participation may not be as strong as that between language and employment or language and earnings, since participation is more a reflection of individual motivation rather than “success”.

The remaining estimation results mirror those of the entire working age population – driven by the fact that 80% of the South African population is black. One important difference in results, however, is observed in primary school attainment – under the narrow definition, the estimated coefficient is now significantly positive, indicating that for black South Africans, any education improves participation. As with our previous findings, we also see a negative change in slope upon use of the broad unemployment definition, yet this is insufficient to deem the effect significantly different from zero.

## **5.2 Employment outcomes**

Results for employment outcomes are presented in Table 6. Using the Heckman and double hurdle models, under both the official and broad definitions of unemployment, we next consider the importance of language for employment outcomes of the entire South African working age population. Columns (1) - (4) provide estimation results for the conventional models, while columns (5) – (8) allow for mother tongue differences. We use the original language dummies: official African languages form the base, to be compared to English, Afrikaans and other (unofficial) languages. We again interact English and Afrikaans with population group to allow for a differing effect of language. Individual marital status is omitted under the presumption that employers are indifferent to the marital status of their employees, yet we retain the household head identifier as a proxy for individual leadership qualities.

Results are equivalent across both Heckman and double hurdle methodologies, and any differences are found only between official and broad employment definitions. The inverse mills ratio for the Heckman model is significant under the narrow definition, but insignificant under the broad definition. This suggests that any sample selection bias is introduced by omitting the disheartened workers, and calls for the use of broad over official definition of unemployed.

The population group dummies are positive across all 8 sets of results. Asian, coloured and white South Africans all appear more likely to be employed than black South Africans under all model specifications. However, once language is taken into account, it becomes clear that the determinants of employment go much deeper than population group alone, with changes in both the marginal effects and the standard errors. The coefficients on the Asian dummies become slightly larger, yet smaller on the coloured and white dummies. Interestingly, the marginal effects now suggest more of a divide in employment propensity of Asian and black than between white and black South Africans, at the base level of all other characteristics.

The marginal effects for English are quite interesting: English improves the employment prospects of black, coloured and white, however, the marginal effect is greatest for black South Africans. In net effect,

Afrikaans- or English-speaking whites remain better off for employment than any other group defined on the basis of population group and language, while the worst off depends on the unemployment definition: under the narrow definition, Asians who speak Afrikaans or English at home are worst off, while under the broad definition, African-speaking blacks fall just short of Afrikaans- or English-speaking Asians.

Males and household heads are more likely to be employed than females and those other than the household head. Age as a proxy for potential experience has a positive, slowly diminishing advantage for employment. Those individuals who have attained primary school level education appear to have a lower propensity for employment than those with no education. Presumably, those who have never been to school would have instead pursued or been forced into work, and hence by default built up vocational skill, leading to employment but not necessarily high levels of income. Returns to secondary and tertiary education are increasingly high in terms of employment outcomes.

Looking to regional disparities, the results provide evidence for rural-urban employment differences, although the direction shifts depending on the definition of the unemployed. Under the official definition, rural dwellers are more likely to be employed than urban dwellers, yet this is reversed under the broad definition. This implies that there are more disheartened workers in the rural areas – in the urban areas there may be greater opportunity to engage in informal self employment in the absence of finding waged work, whereas in the rural areas, the employment opportunities are more restrictive. Finally, all provinces seem to have much higher unemployment rates than the Western Cape, and the underlying country unemployment rate has continued to rise over the three year period.

We next turn to the results for black South Africans in Table 7. Since the majority of black South Africans record African mother tongues, we again define dummies according to prevalence, using English as the base for comparative purposes.

In both selection models and under both unemployment definitions, English mother tongue affords the individual much greater success in employment outcomes than any other language. The models suggest that those who speak Zulu in the home are around 37.3% less likely to be employed under the broad definition than an English mother tongued black South African with the same socio-economic characteristics. And this is the estimated outcome even after controlling for the level of education of the individual. Afrikaans ranks less than English, but greater than other languages for employment. This implies that being able to speak the most widespread language (and even extending this to having more networking potential) is not what is important for employment, but rather, it is one's ability to speak the language of the workplace, this in most cases being English.

### 5.3 Earnings

The two sample selection models provide similar estimates for earnings, as shown in Table 8. With only a difference in how the selectivity correction terms are constructed, there is minimal difference in results, as anticipated. There is some evidence of sample selection bias, and this bias seems to be reduced upon use of the broad definition. The Heckman versions all find significant upward bias in the income equation, while the double hurdle versions are more mixed, tending to attribute bias to the unobservability of employment outcomes for non-participants.

Most relevant to this paper are the results on population group and language. Again, we provide the conventional models in columns (1) – (4) and our language-included alternatives in columns (5) – (8). The results again highlight model misspecification when language is omitted. The conventional models suggest Asian, coloured and white South Africans earn more on average than black South Africans with the same socio-economic characteristics. This population group-based income premium ranges from 10% for coloured and 50% for white South Africans, however, the results change dramatically once language is taken into account, with the premium for being white slashed to almost half that found under the conventional models. Furthermore, our approach deems all other population group dummies insignificant. Such evidence suggests not only that much of the conventional population group effects are largely language effects, but also that, apart from some income differences for the white population, there are no other income differences on the basis of population group. This implies that while some population group-based earnings differentials do still exist, South Africa may be well on its way to breaking down the racial discrimination legacy of its past.

Looking more closely at language particulars, we find those with an English mother tongue to earn an income premium above all other languages, and unlike employment, this premium does not differ according to population group. Despite this population group-constant English language premium, the proportional income improvement for English-speaking white South Africans is minute compared to that experienced by English-speaking black South Africans. Afrikaans is only beneficial for earnings if the individual is Asian, yet the seemingly large marginal effect is not particularly strong and largely netted out by the (albeit insignificant) negative coefficient on Asian.

We again find positive estimated coefficients on male and household head. The models suggest that any education improves earnings, and furthermore, accelerating returns to education. This corroborates the findings of Keswell and Poswell (2002). The Mincerian proxy for experience, age, combined with age<sup>2</sup>, has a positive but slowly diminishing effect on earnings.



Occupations of an agricultural, artisan or operating nature do not seem to yield higher monthly earnings than elementary and domestic occupations, while skilled, and to greater extent managerial/professional occupations, earn higher income. Given that most agricultural and mining operations are set in the rural areas, there are drastically low incomes for workers in this sector, even for the more skilled workers. Monthly earnings in the trade, transport and finance industries are similar to those of the domestic services industry, while manufacturing, utilities and construction sectors do better. Informal sector workers do quite poorly, particularly if they are employees rather than self-employed persons. While this informal disadvantage is largely netted off for the self employed, most informal sector workers would still suffer the low income experienced by workers in elementary occupations.

In addition to lower rural wages, there are also income disparities across provinces and time, even after taking into account regional price differences.

Results are robust when modelling black South Africans alone. Disaggregation of the more prevalent African languages reveal Afrikaans to not only rank below English for earnings, but also below some of the African and unofficial languages. Interestingly, the mother tongue language yielding earnings closest to English is Sepedi, the language with the highest unemployment rate. The most negative partial elasticity is found among those with Sesotho mother tongue, the dominant language in the Free State and Limpopo, and suggests incomes 30.1% lower than those whose mother tongue is English, even after controlling for occupation and education.

## **6. Discussion**

This paper has examined the importance of language for labour force participation, employment and earnings in South Africa. The estimated models suggest that English mother tongue language is important for success in the labour market, even after conditioning on population group and level of educational attainment.

It is recognised that this study has a number of shortcomings and that the significance of the mother tongue variables should be taken with caution. Firstly, information on language proficiency is a missing yet important piece of the puzzle. The individual possessing higher proficiency and thus potentially better communicative skills would find themselves in a better bargaining position for jobs than those who are less well off in their ability to communicate via language. Unfortunately, the October Household Surveys only provide information on the language spoken at home, which we term the mother tongue. No indication is given of ability to speak other languages, nor of their proficiency. Indeed, for a black South African to speak

English at home could imply that the household has had an historically more privileged existence under apartheid than other black South Africans. In this case, English mother tongue could be an indication of class. Data on multiple language proficiency would allow some distinction between a class effect and the degree to which language ability matters in the labour market. Moreover, in attacking the 2001 census, Donnelly (2003) labels responses to a question seeking to identify a single mother tongue from a generalised list as unrevealing. We also must be careful to distinguish the extent to which language determines earnings directly, as opposed to language determining occupation type, which in turn determines earnings.

Further analysis could include interaction effects between combinations of population group, time, education and language. Despite the fact that the South African government spends a large proportion of its budget on schools, it may be that educational attainment matters considerably more for those with African mother tongues. The incorporation of data over a longer time period and population group/time interaction effects may reveal some interesting results concerning population group as a determinant of employment and earnings since apartheid.

Despite the limitations, these tentative results are quite marked: the results suggest that a black South African who speaks English at home is more likely to be employed and to earn a higher income than, for example, a Xhosa speaking black South African with otherwise identical characteristics. Such a result has important implications for individuals which must not be confused with those for policy. While in the current South African context, the individual black South African looking to enhance their employment prospects should prioritise learning the “major” labour force language, English, this advice does not necessarily extend to the government by suggesting adoption of English as the universal official language of South Africa.

Results in the literature based on Hispanic immigrants to the USA would suggest an abolition of multilingualism in favour of one official language. However, South Africa’s situation is uniquely different from that in North America in that the immigrants generally represent a small minority group who also speak a minority language or one that is largely non-existent in the labour force. Yet in South Africa, the dominant (most populous) racial group speak a number of seemingly “minor” languages in the labour force context, while the less populous racial groups speak the “major” labour force language as a result of imbalanced historical factors.

When employees do not share a common lingua franca with their employer and/or fellow employees, the immigration literature screams inefficiency, with uncommunicated comparative skill advantage condemning production to run at an overall average skill level. However, the key is that such comparative skill advantage is *uncommunicated*, not *uncommunicable*. Technological progress need not be thwarted and production need not be inefficient if there was, for example, adequate provision for translators.

Moreover, the racial structure of the South African population is such that efficiency gains can be realised by the minority racial groups opting to embrace the more populous African languages, rather than many Africans striving to learn one language to communicate with the minority population.

The beauty of the new South Africa is that it is all about freedom and equity: the South African constitution embraces freedom of the people through allowing and facilitating each population group to communicate in their own language. It would therefore go against the spirit of the constitution to revert to a single official language, particularly if it were the language of the least populous group. Hence, rather than English as the dominant language in commerce being pushed upon the non-English speaking population, commerce itself could be adapted to embrace the many African languages and subcultures. As a consequence, commerce would then coincide with South Africa's political agenda.

Perhaps also, regional economic development will pave the way for emergence of "dominant languages of commerce" other than English. And these other dominant languages need not be all 11 official languages: one should not ignore the similarities of the African languages, in that proficiency in one African language does not limit the individual to be incommunicable with those of other African tongues. This may even allow for bilingualism, for instance, proficiency in English and (any) one African language for all South Africans. Is this not the approach that is already taken in schools for language education? Indeed, mono- vs. bilingual education is complex and highly sensitive, and our results cannot adequately address the issue. Nonetheless, our findings certainly call for a new way of looking at the South African situation: a new direction for a new, free and equitable South Africa.

## References

- Cragg, J., 1971, "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods", *Econometrica*, 39, 5, 829-844.
- Donnelly, S., 2003, *Language and the Census*, Mail & Guardian, 14/8/03.
- Grenier, G., 1984, "The Effect of Language Characteristics on the Wages of Hispanic American Males", *Journal of Human Resources*, 19, 25-52.
- Heckman, J., 1979, "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 1, 153-162.
- Kingdon, G. and J. Knight, 2004, "Unemployment in South Africa: The Nature of the Beast", *World Development*, 32, 3, 391-408.
- Kingdon, G. and J. Knight, 1999, *Unemployment and Wages in South Africa: A Spatial Approach*, University of Oxford Working Paper WPS/99-12, Oxford.
- Keswell, M. and L. Poswell, 2002, *How Important is Education for Getting Ahead in South Africa?*, CSSR Working Paper No. 22, Cape Town.
- Kossoudji, S., 1988, "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men", *Journal of Labor Economics*, 6, 2, 205-228.
- Maddala, 1983, *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.
- McManus, W., 1985, "Labour Market Costs of Language Disparity. An Interpretation of Hispanic Earnings Differences", *American Economic Review*, 75, 4, 818-827.
- McManus, W., W. Gould and F. Welch, 1983, "Earnings of Hispanic Men: The Role of English Language Proficiency", *Journal of Labour Economics*, 1, 2, 101-130.
- Mora, M., 2003, *An Overview of the Economics of Language in the U.S. Labor Market: Presentation Notes*, American Economic Summer Minority Program Presentation Notes, University of Colorado, Denver.

Rivera-Batiz, F., 1990, "English Language Proficiency and the Economic Progress of Immigrants", *Economics Letters*, 34, 295-300.

Serumaga-Zake, P. and W. Naude, 2003, "Private Rates of Return to Education of Africans in South Africa for 1995: a Double Hurdle Model", *Development Southern Africa*, 20, 4, 515-528.

Southafricainfo, 2004,

[http://www.southafrica.info/ess\\_info/sa\\_glance/demographics/language.htm](http://www.southafrica.info/ess_info/sa_glance/demographics/language.htm).

Smith, 2002, "On Specifying Double Hurdle Models", in *Handbook of Applied Econometrics and Statistical Inference*, A. Ullah, A. T. K. Wan and A. Chaturvedi, (eds), Marcel Dekker, New York.

**Table 1**  
**Labour market outcomes**  
**by official population group classification**

	Unemployment Rate		Average Real Monthly Earnings	
	Official	Expanded	Labour Force Participants	Income Earners
African	28.3	45.4	4.0	11.22
Asian	11.7	15.5	12.2	23.97
Coloured	14.1	21.9	12.6	12.60
White	3.8	5.7	25.3	39.73
Overall	22.6	37.4	6.4	15.70

Source: 1996-1998 October Household Surveys.

**Table 2**  
**Labour market outcomes**  
**by mother tongue**

Mother Tongue Language	Number of Labour Force Participants					Average Real Monthly Earnings	
	African	Asian	Coloured	White	Total	Labour Force Participants	Income Earners
Afrikaans	908	93	13618	6817	21436	1140	1935
English	350	3795	1794	3834	9773	2018	3359
Sepedi	10968	0	17	3	10988	372	1233
Sesotho	12435	2	33	13	12483	402	971
Setswana	12960	0	80	4	13044	444	1119
Xhosa	21455	3	73	35	21566	311	1072
Zulu	27286	12	35	8	27341	390	1162
Other language	13701	147	103	217	14168	431	1222
Total/Overall	100063	4052	15753	10931	130799	631	1570

Source: 1996-1998 October Household Surveys.

**Table 3**  
**Variable definitions**

Variable	Description
Asian Colour White	Dummy variables for population group, taking a value of 1 where the respondent is officially classified as Asian, Coloured and White respectively. Base: African.
Male	Gender dummy taking the value of 1 where the respondent is male. This variable is also interacted with MARRIED to allow a differing effect on labour market participation if the respondent is a husband.
Household head	Dummy variable taking the value of 1 where the respondent is regarded as head of the household.
Primary school Secondary school Diploma Degree Other education	Highest level of educational attainment. Linear regression splines were used to allow differing slopes across primary school, high school, tertiary and other levels of educational attainment. Base is no education.
Rural	Dummy variable taking the value of 1 where the respondent resides in a rural area.
Married	Dummy variable taking the value of 1 where the respondent is married. Used in participation and income equations only. This variable is also interacted with Male to allow a differing effect on labour market participation if the respondent is a husband.
Age Age <sup>2</sup>	Age and Age <sup>2</sup> to allow for a nonlinear effect of age on earnings. This would also capture Mincerian potential experience.
Eastern Cape Northern Cape Free State Kwa-Zulu Natal North West Gauteng MpumaLanga Northern	Dummy variables taking a value of 1 if the individual resides in the named province. Base: Western Cape.
Afrikaans Official African language Unofficial language	Official language dummies. Variables take a value of 1 where the language spoken at home is Afrikaans, one of the African languages deemed an Official language of South Africa and a language not deemed an Official language of South Africa. This set of dummies comprise an alternative to the next set of language dummies. Base: English.
Afrikaans Sepedi Sesotho Setswana Xhosa Zulu Other language	Differentiated African language dummies taking the value of 1 where the language spoken at home is Afrikaans, Sepedi/Northern Sotho, Setswana/Tswana, Isixhosa/Xhosa, Isizulu/Sizulu/Zulu, and a language which is not English and not one of the listed 6. English is the language of particular interest, ranking 8 <sup>th</sup> most common out of all possible responses. Hence, 8 categories were chosen. This set of dummies comprises an alternative to the previous set of language dummies. Base: English.
Managerial occupation Clerical occupation Agricultural occupation	Dummy variable taking the value of 1 if the occupation of the individual is of a managerial, professional or semi-professional; clerical, sales or skilled service; agricultural, artisan or operating nature. Base: Elementary or domestic occupations.
Agricultural industry Construction industry Trade industry	Dummy variable taking the value of 1 if the occupation of the individual is in the industry of agriculture or mining; manufacturing, utilities or construction; trade, transport or finance. Base: domestic services industry.
Informal	A dummy taking the value of 1 for employment in the informal sector. Derived from the main category of occupation and/or, for self-employed persons, an absence of registration of the business for VAT or with the register of companies, the Commissioner of unemployment insurance or the Commissioner of workmen's compensation.
Self employed	A dummy taking the value of 1 if the respondent earned income through a business or other activity of their own. Respondents who also earned income from an employer would also report a 1 for this variable.

**Table 4**  
**Labour Force Participation**  
Double Hurdle and Heckman Sample Selection Models  
All Working Age Adults

Definition of Unemployed	(1)	(2)	(3)	(4)
	Narrow	Broad	Narrow	Broad
Asian	0.040** (4.06)	-0.089** (9.12)	-0.042 (0.69)	-0.153* (2.49)
Coloured	0.160** (23.10)	0.086** (12.96)	0.116** (3.28)	0.091** (2.70)
White	0.064** (9.55)	-0.034** (5.04)	0.090 (1.94)	0.004 (0.09)
Afrikaans			0.127** (6.29)	0.070** (3.61)
Asian x Afrikaans			0.011 (0.12)	0.063 (0.70)
Coloured x Afrikaans			-0.063 (1.57)	-0.063 (1.55)
White x Afrikaans			-0.163** (3.36)	-0.132** (2.58)
English			0.140** (4.00)	0.047 (1.38)
Asian x English			-0.053 (0.75)	0.017 (0.24)
Coloured x English			-0.140** (2.86)	-0.125* (2.40)
White x English			-0.121* (2.12)	-0.050 (0.85)
Unofficial language			-0.017 (0.36)	-0.037 (0.82)
Male	0.135** (38.96)	0.117** (35.71)	0.136** (38.97)	0.117** (35.74)
Household head	0.170** (43.25)	0.120** (30.01)	0.170** (43.25)	0.120** (30.04)
Married	-0.091** (20.14)	-0.143** (31.27)	-0.091** (20.09)	-0.142** (31.13)
Married x Male	0.272** (40.65)	0.306** (47.01)	0.272** (40.64)	0.306** (47.00)
Age	0.085** (111.82)	0.103** (126.78)	0.085** (111.84)	0.103** (126.83)
Age2	-0.001** (98.68)	-0.001** (111.96)	-0.001** (98.70)	-0.001** (112.03)
Primary school	-0.001 (1.46)	-0.005** (5.44)	-0.001 (1.40)	-0.005** (5.41)
Secondary school	0.028** (27.70)	0.019** (19.61)	0.028** (27.80)	0.020** (19.87)
Diploma/certificate	0.239** (28.05)	0.171** (20.93)	0.238** (27.97)	0.171** (20.91)
Degree	0.009 (0.55)	0.020 (1.16)	0.005 (0.29)	0.017 (0.98)
Other education	-0.163** (11.14)	-0.148** (9.85)	-0.161** (11.02)	-0.147** (9.77)
Rural	-0.083** (19.56)	-0.062** (15.41)	-0.082** (19.31)	-0.062** (15.35)
Eastern Cape	-0.150** (19.45)	-0.093** (12.02)	-0.145** (18.67)	-0.090** (11.61)
Northern Cape	-0.033** (3.63)	-0.012 (1.31)	-0.043** (4.67)	-0.020* (2.16)
Free State	0.010 (1.18)	0.015 (1.77)	0.018* (2.00)	0.019* (2.26)
Kwa-Zulu Natal	-0.053** (6.67)	-0.001 (0.10)	-0.046** (5.77)	0.003 (0.40)
North West	-0.042 (4.90)**	0.020* (2.46)	-0.035** (4.00)	0.025** (2.99)
Gauteng	0.029 (3.74)**	0.069** (9.39)	0.035** (4.52)	0.073** (9.80)
Mpumalanga	0.004 (0.47)	-0.011 (1.31)	0.012 (1.39)	-0.006 (0.73)
Northern	-0.120 (13.89)**	-0.098** (11.40)	-0.113** (12.94)	-0.094** (10.82)
y1997	-0.001 (0.13)	0.009 (2.30)*	-0.001 (0.23)	0.009* (2.18)
y1998	0.049 (10.16)**	0.039 (8.52)**	0.049** (10.14)	0.039** (8.51)
Observations	164243	164243	164243	164243

Table reports coefficient estimates; Robust z statistics in parentheses; \* significant at 5%; \*\* significant at 1%



**Table 5**  
**Labour Force Participation**  
Double Hurdle and Heckman Sample Selection Models  
Black South Africans

Definition of Unemployed	(1) Narrow	(2) Broad
Afrikaans	-0.045 (1.14)	0.001 (0.02)
Sepedi	-0.118** (3.49)	-0.057 (1.51)
Sesotho	-0.132** (3.94)	-0.056 (1.49)
Setswana	-0.124** (3.66)	-0.058 (1.56)
Xhosa	-0.117** (3.42)	-0.034 (0.92)
Zulu	-0.102** (2.97)	-0.024 (0.66)
Minority language	-0.088* (2.57)	-0.028 (0.74)
Male	0.117** (30.41)	0.099** (26.05)
Household head	0.144** (33.92)	0.089** (19.30)
Married	-0.078** (16.02)	-0.143** (26.77)
Married x Male	0.251** (33.09)	0.310** (39.64)
Aqe	0.086** (99.97)	0.115** (115.81)
Aqe2	-0.001** (86.17)	-0.001** (100.32)
Primary school	0.004** (4.94)	0.001 (0.71)
Secondary school	0.020** (17.45)	0.010** (8.98)
Diploma/certificate	0.300** (26.73)	0.209** (18.64)
Degree	0.042 (1.59)	0.021 (0.71)
Other education	-0.204** (10.70)	-0.173** (8.24)
Rural	-0.097** (20.49)	-0.077** (16.10)
Eastern Cape	-0.164** (14.80)	-0.144** (11.80)
Northern Cape	0.024 (1.28)	-0.010 (0.52)
Free State	0.016 (1.14)	-0.005 (0.34)
Kwa-Zulu Natal	-0.074** (5.02)	-0.044** (2.82)
North West	-0.036* (2.52)	0.014 (0.91)
Gauteng	0.018 (1.30)	0.049** (3.49)
Mpumalanga	-0.013 (0.90)	-0.045** (2.91)
Northern	-0.120** (8.24)	-0.115** (7.22)
y1997	0.005 (1.10)	0.020** (4.08)
y1998	0.067** (12.32)	0.056** (10.44)
Observations	126838	126838

Table reports coefficient estimates; Robust z statistics in parentheses; \* significant at 5%; \*\* significant at 1%

**Table 6**  
**Employment**  
All Working Age Adults

Model Specification Definition of Unemployed	(1) Heckman		(3) Double Hurdle		(5) Heckman		(7) Double Hurdle	
	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad
Asian	0.125** (16.12)	0.237** (22.63)	0.126** (16.49)	0.238** (22.54)	0.174** (4.27)	0.277** (5.22)	0.173** (4.26)	0.276** (5.21)
Coloured	0.109** (16.91)	0.201** (28.26)	0.117** (20.40)	0.201** (27.54)	0.067* (1.98)	0.116** (3.23)	0.074* (2.26)	0.117** (3.26)
White	0.177** (31.37)	0.301** (39.67)	0.178** (32.17)	0.301** (39.70)	0.105* (2.54)	0.219** (4.14)	0.110** (2.71)	0.219** (4.15)
Afrikaans					0.074** (3.85)	0.146** (6.61)	0.083** (4.42)	0.147** (6.66)
Asian x Afrikaans					-0.378** (2.66)	-0.285* (2.32)	-0.375** (2.65)	-0.284* (2.31)
Coloured x Afrikaans					-0.025 (0.56)	-0.044 (0.94)	-0.030 (0.69)	-0.045 (0.96)
White x Afrikaans					0.062 (1.25)	-0.003 (0.05)	0.048 (0.95)	-0.006 (0.08)**
English					0.150** (6.32)	0.262** (7.93)	0.154** (6.60)	0.263 (7.94)
Asian x English					-0.546** (4.39)	-0.454** (4.86)	-0.550** (4.44)	-0.454** (4.86)
Coloured x English					-0.126* (2.06)	-0.187** (2.89)	-0.140* (2.26)	-0.188** (2.91)
White x English					-0.070 (1.02)	-0.170* (2.02)	-0.083 (1.20)	-0.171* (2.03)
Unofficial language					0.072 (1.85)	0.091 (1.75)	0.070 (1.81)	0.090 (1.73)
Male	0.021** (3.89)	0.091** (25.05)	0.039** (12.22)	0.089** (16.43)	0.020** (3.66)	0.088** (16.25)	0.039** (12.17)	0.091** (25.01)
Household head	0.130** (24.38)	0.213** (53.28)	0.146** (41.29)	0.211** (43.51)	0.129** (24.20)	0.211** (43.37)	0.146** (41.31)	0.213** (53.28)
Age	0.012** (5.78)	0.028** (28.54)	0.020** (24.31)	0.026** (10.96)	0.012** (5.57)	0.026** (10.78)	0.020** (24.48)	0.028** (28.71)
Age2	-0.000** (3.30)	-0.000** (19.49)	-0.000** (16.85)	-0.000** (7.60)	-0.000** (3.11)	-0.000** (7.44)	-0.000** (17.00)	-0.000** (19.64)
Primary school	-0.006** (6.51)	-0.002** (2.63)	-0.006** (6.71)	-0.002* (2.55)	-0.005** (6.40)	-0.002* (2.37)	-0.006** (6.61)	-0.002* (2.47)
Secondary school	0.007** (6.06)	0.019** (17.09)	0.009** (9.84)	0.019** (15.91)	0.006** (5.47)	0.018** (15.30)	0.009** (9.36)	0.019** (16.56)
Diploma/certificate	0.108** (15.33)	0.220** (26.98)	0.118** (19.75)	0.219** (25.55)	0.107** (15.14)	0.218** (25.33)	0.118** (19.66)	0.219** (26.81)
Degree	0.025 (1.68)	0.023 (1.11)	0.024 (1.61)	0.023 (1.11)	0.022 (1.48)	0.018 (0.84)	0.021 (1.39)	0.018 (0.84)
Other education	-0.084** (5.44)	-0.149** (7.99)	-0.095** (6.27)	-0.148** (7.85)	-0.081** (5.26)	-0.145** (7.67)	-0.093** (6.13)	-0.147** (7.84)
Rural	0.033** (7.55)	-0.011* (2.41)	0.026** (6.31)	-0.011* (2.20)	0.035** (7.97)	-0.008 (1.73)	0.027** (6.65)	-0.009* (1.98)
Eastern Cape	-0.156** (15.00)	-0.265** (26.01)	-0.175** (17.98)	-0.264** (25.41)	-0.149** (14.45)	-0.255** (24.40)	-0.168** (17.33)	-0.256** (24.98)
Northern Cape	-0.050** (4.81)	-0.093** (7.91)	-0.053** (5.13)	-0.093** (7.89)	-0.054** (5.11)	-0.101** (8.50)	-0.058** (5.56)	-0.101** (8.53)
Free State	-0.039** (4.20)	-0.059** (5.45)	-0.039** (4.12)	-0.059** (5.46)	-0.034** (3.65)	-0.049** (4.55)	-0.033** (3.49)	-0.049** (4.53)
Kwa-Zulu Natal	-0.076** (8.23)	-0.150** (14.75)	-0.081** (8.87)	-0.150** (14.75)	-0.070** (7.65)	-0.140** (13.70)	-0.075** (8.23)	-0.140** (13.68)
North West	-0.046** (4.77)	-0.124** (11.49)	-0.050** (5.22)	-0.124** (11.51)	-0.040** (4.21)	-0.114** (10.54)	-0.044** (4.59)	-0.114** (10.51)
Gauteng	-0.078** (9.13)	-0.113** (11.71)	-0.075** (8.86)	-0.114** (11.71)	-0.073** (8.57)	-0.106** (10.77)	-0.070** (8.22)	-0.105** (10.73)
Mpumalanga	-0.061** (6.25)	-0.063** (5.72)	-0.060** (6.21)	-0.063** (5.70)	-0.055** (5.68)	-0.052** (4.72)	-0.054** (5.54)	-0.052** (4.74)
Northern	-0.140** (12.28)	-0.210** (18.00)	-0.154** (13.88)	-0.209** (17.66)	-0.132** (11.75)	-0.198** (16.70)	-0.146** (13.29)	-0.199** (17.06)
y1997	-0.012** (2.66)	-0.017** (3.42)	-0.012** (2.72)	-0.017** (3.43)	-0.011** (2.60)	-0.017** (3.38)	-0.012** (2.67)	-0.017** (3.36)
y1998	-0.048** (9.91)	-0.026** (4.75)	-0.044** (9.29)	-0.026** (4.77)	-0.048** (9.93)	-0.026** (4.72)	-0.044** (9.25)	-0.025** (4.67)
$\lambda_p$	-0.060** (4.21)			-0.009 (0.58)	-0.064** (4.48)	-0.012 (0.80)		
Observations	75992	94001	75992	94001	75992	94001	75992	94001

Table reports coefficient estimates; Robust z statistics in parentheses; \* significant at 5%; \*\* significant at 1%;  $\lambda_p$  is the inverse Mills ratio from the participation equation.

**Table 7**  
**Employment**  
**Black South Africans**

Model Specification Definition of Unemployed	(1) Heckman		(3) Double Hurdle	
	(2) Broad	(4) Broad	Narrow	Broad
Afrikaans	-0.193** (3.35)	-0.233** (4.24)	-0.195** (3.36)	-0.233** (4.24)
Sepedi	-0.321** (6.26)	-0.357** (7.55)	-0.327** (6.38)	-0.357** (7.56)
Sesotho	-0.309** (6.13)	-0.367** (7.77)	-0.316** (6.27)	-0.367** (7.77)
Setswana	-0.294** (5.84)	-0.350** (7.35)	-0.300** (5.97)	-0.350** (7.35)
Xhosa	-0.308** (6.23)	-0.364** (7.64)	-0.313** (6.34)	-0.364** (7.64)
Zulu	-0.312** (6.54)	-0.373** (7.81)	-0.316** (6.63)	-0.373** (7.81)
Minority language	-0.282** (5.64)	-0.320** (6.64)	-0.286** (5.70)	-0.320** (6.64)
Male	0.050** (6.78)	0.113** (18.23)	0.058** (13.27)	0.113** (25.88)
Household head	0.186** (26.43)	0.247** (45.58)	0.193** (40.60)	0.247** (52.17)
Age	0.021** (6.67)	0.031** (9.76)	0.026** (21.84)	0.032** (25.83)
Age2	-0.000** (4.59)	-0.000** (6.88)	-0.000** (15.03)	-0.000** (17.63)
Primary school	-0.005** (4.26)	0.001 (0.46)	-0.005** (4.12)	0.001 (0.46)
Secondary school	0.007** (4.69)	0.017** (11.94)	0.008** (6.00)	0.017** (12.19)
Diploma/certificate	0.162** (15.35)	0.289** (25.86)	0.168** (19.86)	0.289** (27.57)
Degree	0.059* (2.57)	0.086** (3.08)	0.059* (2.58)	0.086** (3.08)
Other education	-0.133** (5.98)	-0.208** (8.76)	-0.141** (6.48)	-0.209** (8.86)
Rural	0.018** (2.79)	-0.037** (6.13)	0.013* (2.31)	-0.037** (6.38)
Eastern Cape	-0.170** (10.19)	-0.222** (15.35)	-0.180** (11.60)	-0.223** (15.75)
Northern Cape	-0.015 (0.69)	0.028 (1.22)	-0.014 (0.64)	0.028 (1.22)
Free State	-0.014 (0.84)	0.011 (0.65)	-0.014 (0.80)	0.011 (0.65)
Kwa-Zulu Natal	-0.036* (1.97)	-0.077** (4.26)	-0.039* (2.18)	-0.077** (4.26)
North West	-0.025 (1.41)	-0.065** (3.70)	-0.027 (1.51)	-0.065** (3.70)
Gauteng	-0.059** (3.66)	-0.059** (3.68)	-0.059** (3.63)	-0.059** (3.68)
Mpumalanga	-0.038* (2.10)	-0.011 (0.62)	-0.038* (2.14)	-0.011 (0.63)
Northern	-0.132** (6.53)	-0.159** (8.38)	-0.139** (6.97)	-0.159** (8.45)
y1997	-0.011 (1.80)	-0.016** (2.70)	-0.011 (1.79)	-0.016** (2.70)
y1998	-0.058** (8.44)	-0.020** (2.96)	-0.055** (8.37)	-0.020** (3.01)
$\lambda_p$	-0.031 (1.49)	-0.001 (0.06)		
Observations	52672	69242	52672	69242

Table reports coefficient estimates; Robust z statistics in parentheses; \* significant at 5%; \*\* significant at 1%;  $\lambda_p$  is the inverse Mills ratio from the participation equation.

**Table 8**  
**Income**  
All Working Age Adults

Model Specification	(1) Heckman		(3) Double Hurdle		(5) Heckman		(7) Double Hurdle	
	Narrow	Broad	Narrow	Broad	Narrow	Broad	Narrow	Broad
Asian	0.133** (5.54)	0.133** (5.37)	0.157** (6.45)	0.154** (6.07)	-0.229 (1.30)	-0.225 (1.28)	-0.181 (1.04)	-0.198 (1.13)
Coloured	0.108** (6.46)	0.109** (6.34)	0.102** (6.01)	0.108** (6.30)	-0.005 (0.07)	-0.005 (0.06)	-0.013 (0.17)	-0.010 (0.13)
White	0.487** (26.58)	0.492** (27.14)	0.518** (27.17)	0.509** (27.24)	0.271* (2.28)	0.267* (2.25)	0.271* (2.30)	0.271* (2.29)
Afrikaans					0.002 (0.05)	-0.000 (0.01)	-0.007 (0.16)	-0.001 (0.03)
Asian x Afrikaans					0.481* (2.06)	0.477* (2.04)	0.454 (1.95)	0.463* (1.99)
Coloured x Afrikaans					0.084 (0.94)	0.084 (0.94)	0.091 (1.01)	0.088 (0.98)
White x Afrikaans					0.215 (1.68)	0.224 (1.75)	0.261* (2.06)	0.241 (1.90)
English					0.221** (4.43)	0.223** (4.47)	0.234** (4.69)	0.231** (4.63)
Asian x English					0.160 (0.87)	0.151 (0.83)	0.123 (0.68)	0.138 (0.76)
Coloured x English					0.086 (0.91)	0.085 (0.90)	0.094 (1.00)	0.091 (0.97)
White x English					0.030 (0.23)	0.034 (0.26)	0.043 (0.33)	0.036 (0.28)
Unofficial language					0.143 (1.16)	0.147 (1.19)	0.166 (1.37)	0.158 (1.30)
Male	0.371** (42.23)	0.367** (40.02)	0.323** (26.60)	0.342** (30.06)	0.372** (42.42)	0.368** (40.20)	0.322** (26.47)	0.341** (29.98)
Household head	0.089** (6.77)	0.093** (7.19)	0.071** (5.12)	0.082** (6.19)	0.088** (6.81)	0.090** (7.05)	0.069** (5.05)	0.079** (5.99)
Age	0.061** (22.48)	0.062** (23.53)	0.042** (9.88)	0.048** (10.50)	0.061** (22.58)	0.061** (23.47)	0.041** (9.70)	0.047** (10.29)
Age2	-0.001** (19.97)	-0.001** (20.79)	-0.000** (7.70)	-0.000** (8.36)	-0.001** (20.06)	-0.001** (20.78)	-0.000** (7.54)	-0.000** (8.19)
Primary school	0.047** (20.81)	0.046** (20.69)	0.045** (20.22)	0.046** (20.88)	0.047** (21.06)	0.046** (20.94)	0.046** (20.45)	0.047** (21.15)
Secondary school	0.101** (36.23)	0.100** (35.15)	0.096** (32.96)	0.099** (34.25)	0.099** (35.50)	0.098** (34.38)	0.093** (32.04)	0.096** (33.42)
Diploma/certificate	0.291** (18.69)	0.291** (18.07)	0.270** (16.64)	0.280** (17.17)	0.290** (18.80)	0.288** (18.07)	0.267** (16.63)	0.277** (17.11)
Degree	0.159** (7.60)	0.160** (7.66)	0.159** (7.63)	0.158** (7.59)	0.156** (7.47)	0.157** (7.55)	0.157** (7.50)	0.156** (7.47)
Other education	-0.409** (11.33)	-0.410** (11.35)	-0.391** (10.81)	-0.399** (11.04)	-0.407** (11.34)	-0.408** (11.35)	-0.388** (10.79)	-0.396** (11.02)
Managerial occupation	0.371** (16.29)	0.371** (16.27)	0.371** (16.29)	0.372** (16.31)	0.365** (15.98)	0.364** (15.96)	0.365** (15.99)	0.365** (16.00)
Clerical occupation	0.048* (2.24)	0.048* (2.23)	0.047* (2.19)	0.048* (2.22)	0.044* (2.04)	0.044* (2.03)	0.043* (1.99)	0.044* (2.03)
Agricultural occupation	-0.002 (0.09)	-0.002 (0.10)	-0.003 (0.16)	-0.002 (0.11)	-0.003 (0.17)	-0.004 (0.17)	-0.005 (0.24)	-0.004 (0.19)
Agricultural industry	-0.197** (14.37)	-0.198** (14.41)	-0.198** (14.43)	-0.197** (14.36)	-0.198** (14.42)	-0.199** (14.46)	-0.199** (14.49)	-0.198** (14.42)
Construction industry	0.111** (9.82)	0.111** (9.81)	0.111** (9.77)	0.111** (9.80)	0.108** (9.53)	0.108** (9.53)	0.108** (9.48)	0.108** (9.52)
Trade industry	-0.014 (1.37)	-0.014 (1.37)	-0.015 (1.46)	-0.015 (1.41)	-0.019 (1.84)	-0.019 (1.84)	-0.020 (1.94)	-0.020 (1.89)
Formal sector	0.428** (21.81)	0.429** (21.81)	0.428** (21.79)	0.428** (21.82)	0.427** (21.68)	0.428** (21.68)	0.427** (21.67)	0.427** (21.69)
Self employed	0.319** (9.69)	0.320** (9.70)	0.321** (9.75)	0.320** (9.72)	0.314** (9.51)	0.315** (9.53)	0.316** (9.58)	0.315** (9.55)
Rural	-0.326** (28.79)	-0.319** (28.26)	-0.301** (24.87)	-0.312** (27.40)	-0.322** (28.36)	-0.314** (27.81)	-0.295** (24.35)	-0.307** (26.91)
Eastern Cape	-0.138** (6.89)	-0.137** (6.69)	-0.127** (6.28)	-0.134** (6.56)	-0.136** (6.90)	-0.133** (6.60)	-0.125** (6.26)	-0.130** (6.46)
Northern Cape	-0.206** (10.42)	-0.206** (10.39)	-0.206** (10.40)	-0.206** (10.41)	-0.193** (9.62)	-0.192** (9.54)	-0.191** (9.50)	-0.192** (9.53)
Free State	-0.384** (19.82)	-0.384** (19.85)	-0.390** (20.09)	-0.387** (19.95)	-0.382** (19.58)	-0.383** (19.60)	-0.389** (19.88)	-0.385** (19.71)
Kwa-Zulu Natal	0.032 (1.76)	0.034 (1.87)	0.033 (1.84)	0.030 (1.65)	0.027 (1.52)	0.031 (1.68)	0.028 (1.56)	0.026 (1.43)
North West	0.019 (1.01)	0.025 (1.29)	0.025 (1.28)	0.020 (1.03)	0.020 (1.02)	0.026 (1.33)	0.024 (1.27)	0.021 (1.05)
Gauteng	0.097** (6.24)	0.096** (6.17)	0.081** (5.09)	0.086** (5.45)	0.097** (6.18)	0.096** (6.15)	0.079** (4.96)	0.085** (5.39)
Mpumalanga	-0.021 (1.07)	-0.025 (1.31)	-0.030 (1.53)	-0.025 (1.28)	-0.021 (1.07)	-0.025 (1.29)	-0.031 (1.58)	-0.025 (1.27)
Northern	-0.045* (1.99)	-0.047* (2.10)	-0.039 (1.73)	-0.042 (1.87)	-0.045* (1.97)	-0.046* (2.03)	-0.039 (1.72)	-0.040 (1.79)
y1997	-0.025* (2.33)	-0.025* (2.35)	-0.026* (2.43)	-0.027* (2.48)	-0.023* (2.17)	-0.023* (2.17)	-0.024* (2.26)	-0.025* (2.31)
y1998	-0.018 (1.53)	-0.024* (2.07)	-0.036** (2.90)	-0.029* (2.48)	-0.019 (1.58)	-0.025* (2.11)	-0.037** (3.01)	-0.030* (2.54)
$\lambda_p$			-0.196** (5.60)	-0.118** (3.54)			-0.204** (5.83)	-0.124** (3.70)

$\lambda_e^{HK}$	-0.154** (3.25)	-0.105** (2.91)			-0.157** (3.42)	-0.113** (3.20)		
$\lambda_e^{DH}$			-0.011 (0.21)	-0.074* (1.99)			-0.008 (0.16)	-0.081* (2.20)
Constant	-0.224** (3.02)	-0.240** (3.18)	0.239* (2.14)	0.075 (0.67)	-0.220** (2.99)	-0.225** (2.99)	0.265* (2.37)	0.106 (0.94)
Observations	49538	49538	49538	49538	49538	49538	49538	49538

Table reports coefficient estimates; Robust z statistics in parentheses; \* significant at 5%; \*\* significant at 1%;  $\lambda_p$  is the inverse Mills ratio from the participation equation, while  $\lambda_e^{HK}$  is that from the Heckman sample selection-corrected employment equation and  $\lambda_e^{DH}$  its Double Hurdle counterpart.

**Table 9**  
**Income**  
Black South Africans

Model Specification Definition of Unemployed	(1)	(2)	(3)	(4)
	Heckman		Double Hurdle	
	Narrow	Broad	Narrow	Broad
Afrikaans	-0.192** (2.83)	-0.195** (2.88)	-0.218** (3.22)	-0.211** (3.11)
Sepedi	-0.127* (2.27)	-0.133* (2.39)	-0.159** (2.84)	-0.156** (2.79)
Sesotho	-0.243** (4.41)	-0.244** (4.40)	-0.268** (4.85)	-0.269** (4.82)
Setswana	-0.234** (4.27)	-0.236** (4.29)	-0.256** (4.68)	-0.257** (4.66)
Xhosa	-0.143** (2.61)	-0.147** (2.68)	-0.176** (3.20)	-0.174** (3.16)
Zulu	-0.157** (2.88)	-0.164** (3.02)	-0.199** (3.64)	-0.193** (3.53)
Minority language	-0.138* (2.56)	-0.146** (2.71)	-0.174** (3.20)	-0.168** (3.11)
Male	0.382** (29.86)	0.372** (24.65)	0.322** (20.76)	0.351** (22.07)
Household head	0.000 (0.02)	0.006 (0.23)	0.004 (0.14)	0.018 (0.67)
Age	0.052** (10.84)	0.054** (11.97)	0.024** (3.72)	0.036** (5.43)
Age2	-0.001** (10.46)	-0.001** (11.57)	-0.000* (2.46)	-0.000** (4.08)
Primary school	0.047** (18.79)	0.046** (18.48)	0.044** (17.05)	0.046** (18.46)
Secondary school	0.091** (26.61)	0.089** (24.48)	0.086** (24.40)	0.090** (24.53)
Diploma/certificate	0.323** (10.75)	0.316** (9.55)	0.299** (9.76)	0.322** (9.71)
Degree	0.211** (7.05)	0.213** (7.12)	0.219** (7.32)	0.217** (7.26)
Other education	-0.479** (9.68)	-0.478** (9.60)	-0.458** (9.22)	-0.475** (9.55)
Managerial occupation	0.420** (14.91)	0.420** (14.88)	0.419** (14.88)	0.420** (14.91)
Clerical occupation	0.057* (2.19)	0.056* (2.17)	0.054* (2.09)	0.055* (2.12)
Agricultural occupation	0.021 (0.87)	0.020 (0.85)	0.018 (0.77)	0.020 (0.83)
Agricultural industry	-0.185** (11.05)	-0.185** (11.10)	-0.184** (11.03)	-0.184** (11.01)
Construction industry	0.111 (7.50)**	0.110** (7.48)	0.110** (7.46)	0.110** (7.48)
Trade industry	-0.015 (1.12)	-0.016 (1.15)	-0.017 (1.24)	-0.017 (1.21)
Formal sector	0.398** (17.42)	0.398** (17.41)	0.397** (17.41)	0.398** (17.44)
Self employed	0.211** (5.20)	0.212** (5.23)	0.215** (5.29)	0.214** (5.27)
Rural	-0.294** (22.69)	-0.278** (20.90)	-0.252** (17.64)	-0.271** (20.23)
Eastern Cape	-0.121** (3.42)	-0.120** (3.34)	-0.107** (3.01)	-0.125** (3.46)
Northern Cape	-0.173** (4.23)	-0.186** (4.52)	-0.184** (4.49)	-0.180** (4.36)
Free State	-0.344** (10.31)	-0.352** (10.57)	-0.354** (10.62)	-0.349** (10.50)
Kwa-Zulu Natal	0.055 (1.56)	0.062 (1.76)	0.073* (2.08)	0.060 (1.70)
North West	0.064 (1.92)	0.072* (2.14)	0.070* (2.11)	0.063 (1.88)
Gauteng	0.128** (4.16)	0.123** (4.02)	0.105** (3.42)	0.111** (3.62)
Mpumalanga	-0.031 (0.92)	-0.042 (1.25)	-0.038 (1.12)	-0.036 (1.06)
Northern	-0.059 (1.50)	-0.063 (1.61)	-0.053 (1.37)	-0.063 (1.62)
y1997	-0.027 (1.98)*	-0.027 (1.94)	-0.031* (2.25)	-0.032* (2.27)
y1998	-0.019 (1.18)	-0.035 (2.34)*	-0.059** (3.42)	-0.045** (2.98)
$\lambda_p$			-0.295** (6.51)	-0.163** (3.67)
$\lambda_c^{HK}$	-0.304** (3.32)	-0.221** (3.10)		
$\lambda_c^{DH}$			-0.037 (0.36)	-0.126 (1.66)
Constant	0.203 (1.44)	0.190 (1.30)	0.893** (5.04)	0.574** (3.29)
Observations	32190	32190	32190	32190

Table reports coefficient estimates; Robust z statistics in parentheses; \* significant at 5%; \*\* significant at 1%;  $\lambda_p$  is the inverse Mills ratio from the participation equation, while  $\lambda_c^{HK}$  is that from the Heckman sample selection-corrected employment equation and  $\lambda_c^{DH}$  its Double Hurdle counterpart.