

Language and Labour in South Africa

A new approach for a new South Africa

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

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 180,000 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 race. After conditioning on a number of socio-economic and demographic factors, we find the English language to be 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.

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 race 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 race, 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 race no longer determines one’s fate. Disappointingly, studies continue to find race 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 race 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 race into what is of importance to an individual's relative success in the labour market. Is it race 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: Cragg's (1971) double hurdle and Heckman's (1979) sample selection model, with mother tongue included in addition to race 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 languages¹. 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.

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

Literature on the economics of language in the labour market is limited. The majority involve consideration of the role of language on 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, our sample data indicates that English mother tongue adds an average earnings premium in the order of 200% over the most widely spoken African language, Zulu (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 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, β^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.

Sequential double hurdle model

We assume that the employment decision is subsequent to the participation decision. In this case y_i^I is observed only if $I_i^P = 1$ and $I_i^E = 1$, and we have a sequential model whereby employment is independent of participation. This independence assumption is somewhat dubious, however, since it is not likely to make considerable difference to the results, we leave it for future work. For the model specification relaxing this assumption, the reader is referred to Maddala (1993).

Assuming normality of the error terms, the sequential 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

In estimation, Heckman's model differs from the sequential double hurdle model in its inclusion of the two correction factors. 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. 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 λ_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 188,985 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 monthly income and deduced from waged and/or self-employed sources.

Participation and employment are the secondary binary dependent variables. 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 Surveys from 1996 onwards 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 the different aspects of labour force participation, employment and earnings. We also condition on a number of socio-economic and demographic variables including population group, gender, household head, education level, rural residency, marital status, age and time dummies. Linear regression splines for education level are used to allow differing slopes across lower and upper primary, lower and upper secondary, tertiary and other levels of educational attainment. For the income equation, we also include hours worked and a dummy variable for employment in the informal sector. The married dummy is omitted from the employment equation 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. The move from the official to broad definition means that some 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 if given the chance, but are not actively seeking work. While the move from official to broad does not alter the qualitative directions of the results, some variables become significant upon use of the broad definition. Note that the estimated sequential double hurdle and Heckman participation models are identical in this first stage.

All of the race dummies are significant in the participation models even after controlling for language. We find that coloured South Africans are more likely to participate, while Asian and white South Africans are significantly less likely to participate in the labour market than are blacks. This could be reflective of the “old money” syndrome of whites: white South Africans have historical wealth as a safety net in harder times, enabling them to drop out of the labour force much more easily than the invariably poorer black South African.

As one would expect, males, household heads and married persons are much more likely to participate than their respective base counterparts under the broad definition. This result highlights the importance of the household as an economic and cultural unit in South Africa: the decision to participate in the labour force is made in conjunction with the cultural obligations of the extended family. Under the narrower definition of unemployed, marriage does not appear to be statistically significant.

Educational attainment is important for labour force participation. Those with Standard 1 education are more likely to enter the labour force than those with no education. The slope eases off for those with Standard 2 through to Standard 4, yet remains marginally steeper than the base of no education. The direction is reversed for those in lower secondary school (Standards 5 to 7). This could reflect two opposing effects. These people may comprise the discouraged workers who do not enter the labour force because they lack a sufficient educational standard to find work. With the model defined over the entire working age population, these individuals could also be those currently in school and hence not participating in the labour force at this stage. Standard 8 and above education significantly

improves the likelihood that the individual will join the labour force. This reflects the fact that these people perceive themselves as more employable; perhaps schooling has improved availability of information about employment networks and prospects.

The model provides some quite interesting insight into language as a determinant of labour force participation. Results under the broad definition suggest that when the language spoken at home is any language other than English, the individual is less likely to participate in the labour force. Indeed, it seems that language is one of the key factors in labour market participation under this definition. When one cannot communicate in their mother tongue, doors close and networks cannot be formed, forcing even those willing to work to become discouraged. Of the eight language categories, it is two of the common African languages, Xhosa and Sepedi, which record extremely large negative marginal effects.

Mother tongues of Zulu, Sesotho and Setswana become insignificant upon narrowing the unemployed to the official definition. One could deduce from this result that those who speak these particular languages at home are the discouraged rather than the disheartened workers: they are forced out of the labour market altogether and are not just simply not looking for work.

Finally, while the labour force has expanded over the three years, the participation rate of rural dwellers is lower than the urban.

5.2 Employment outcomes

Results for employment outcomes are presented in Tables 5 and 6. There is little difference between the double hurdle and Heckman models under this specification. We note that the Heckman model's sample selection correction term is insignificant under the official definition, but highly significant under the broad definition.

The unemployment rate appears to have risen over the three year period. Race dummies are again significant, with Asian, coloured and white South Africans all more likely to be employed than black South Africans. Males and household heads are more likely to be employed than females and those other than the household head. Interestingly, under the official unemployment definition, rural dwellers are more likely to be employed than urban dwellers, yet this result is reversed under the broad definition.

Age as a proxy for potential experience has a diminishing effect on the probability of employment. Educational attainment has a mixed effect on employment outcomes. Under the official definition, those with Standard 1 education are actually less likely to be employed than those with no education. Under the broad definition, however, education holds no advantage in employment prospects until secondary school level. In fact, the Heckman model finds lower secondary schooling (Standards 5-7) to reduce the chances of employment, indicating that vocational skill may be more important than these levels of education. Returns from upper secondary and tertiary education have a large positive effect on the probability of employment. These results reconcile with Keswell and Poswell (2002) who find that returns to education accelerate rather than diminish. These results are not surprising, considering that the government has injected substantial funds into improving education. In fact, South Africa can boast school enrolment ratios higher than other countries in the developing world.

In both selection models and under both unemployment definitions, English mother tongue affords the individual much greater success in employment outcomes. The model suggests that those who speak one of the five listed African languages at home are somewhere in the order of 20% less likely to be employed than an English mother tongued individual with the same socio-economic characteristics. And this is the estimated outcome even after controlling for race and the level of education of the individual.

5.3 Earnings

The two sample selection models provide similar estimates for earnings, as shown in Tables 7 and 8. All the appropriate sample selection correction terms are significant.

We again observe positive coefficients on male, household head and married dummies. The models suggest that any education improves earnings. While Standards 1 through to 7 provide similar returns to education, completion of secondary education, and even more so, tertiary education, greatly improve the earnings potential of the individual South African. The Mincerian proxy for experience, age, combined with age², has a positive but diminishing effect on earnings. While significant, hours worked has little partial effect on income, owing mainly to the small variation in hours worked for individuals in the sample. Informal sector and rural workers have lower incomes than those in formal jobs and the urban area respectively.

Most relevant to this paper are the results on race and language. White South Africans appear to earn a premium over the other racial groups, yet the models consistently suggest that incomes of Asians and coloured South Africans are insignificantly different to blacks with the same socio-economic and demographic characteristics. This implies that while earnings differentials do still exist, South Africa may be well on its way to breaking down discrimination on the basis of race alone. English mother tongue provides an income premium above all the other languages. Sesotho, the dominant language in the Free State and Limpopo, records the lowest partial elasticity, suggesting that incomes are 40% less than those whose mother tongue is English.

The Afrikaans dummy has a surprisingly large negative estimated semi-elasticity. Notably, Afrikaans speakers are drawn from a wide range of racial backgrounds, the majority coloured and white. Indeed, the effect on income for a coloured South African to speak Afrikaans speaker could well be different to that of a white South African. Preliminary study into the interaction of colour and Afrikaans dummy variables has been undertaken with some interesting results. For the additive model in this paper, we find the coloured dummy variable to be insignificant, yet when this effect is interacted with Afrikaans, we find significant race and language effects. Briefly, our results indicate that being a coloured South African with an Afrikaans mother tongue has a greater detrimental effect on income than having an Afrikaans mother tongue and being a white South African. One explanation could be that Afrikaans speaking coloureds may be concentrated in particular low-wage occupations.

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 race and level of educational attainment.

However, it is recognised that this study has a number of shortcomings. 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 whom 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 fortunate existence than other black South Africans under apartheid. 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.

The significance of the mother tongue dummies should be taken with caution. We have not conditioned on magisterial district in our model. It could well be that language is proxying for area, whereby the regions where particular languages are spoken are in fact the poorer areas of South Africa with higher unemployment rates. Indeed, the segregation of white and black South Africans during apartheid meant that whole geographical areas were designated according to their perceived political status. Similarly, we 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 race, 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 more data and a race/time interaction effect may reveal some interesting results concerning race as a determinant of employment and earnings since apartheid. Finally, we recognise that the decision to join the queue for a job is not independent of the probability of the individual finding herself employed. Accordingly, our model specification should be altered to accommodate the resulting correlation in the errors.

Despite the limitations, these tentative results are quite marked: a black South African with no education who speaks English at home (albeit paradoxical) is modelled as more likely to participate in the labour force, more likely to be employed, and predicted to earn a higher income than, for example, a Xhosa speaking African with otherwise identical characteristics. Such a result has important implications for policy concerning South Africa's multilingual political and societal stance.

The policy directive flowing from the immigration literature would point in the direction of abolition of multilingualism in favour of one official language: English. However, South

Africa's situation is unique, and not simply from a moral or ethical viewpoint. In the immigration literature, immigrants represent a small minority group speaking a minor language. In South Africa, we have the dominant (most populous) group speaking a number of seemingly "minor" languages, while the less populous group speaks the "major" language as a result of imbalanced historical factors. The message of this paper is that, from the point of view of the individual black South African looking to enhance their employment prospects, their priority ought to be to learn the "major" language, English. However, this does not necessarily extend to the government by suggesting adoption of English as the universal official language of South Africa. The new South Africa 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 minority (least populous) group.

The immigrant/minority language literature suggests that technological progress is thwarted by lack of communication via language between workers, and moreover, since comparative skill advantages cannot be communicated, production results at an overall average rate of skill level rather than benefiting from specialisation according to skill. However, multilingualism is not always costly (Coulmas, 1992). Consider the case of multilingual communication via a translator, for example. Efficiency gains can be realised by minority language groups learning the more populous African languages, rather than many Africans learning one language to communicate with the minority population. 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 African language and subculture. Consequently, commerce would then coincide with South Africa's political agenda.

"What seems to be needed is a progressive race narrative that is able to challenge the neo-liberal war on the poor without abandoning the need for blacks to be the authors of their own destiny. It is for the historically dominant bodies to learn to listen, empathise and follow, without crowding out the voices of the marginalised. To do otherwise is to turn solidarity into imperialism."

Mngxitama, 22/6/04.

References

- Coulmas, F., 1992, *Language and Economy*, Blackwell Publishers, Oxford.
- 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.
- 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.
- Mincer, J., 1974, *Schooling, Experience, and Earnings*. New York: Columbia University Press.

- Mngxitama, A., 2004, *Let Black Voices Speak for the Voiceless*, Mail & Guardian, 22/6/04.
- 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.

Table 1
Average sample unemployment rates and average monthly earnings
by official race classification

	Unemployment rate		Average monthly earnings
	Official	Expanded	
African	28.22	45.40	1582
Asian	11.70	15.51	3218
Coloured	14.12	21.86	1736
White	3.80	5.67	5674
Overall	22.56	37.40	2213

Source: 1996-1998 October Household Surveys.

Table 2
Number of respondents and average monthly earnings
by mother tongue

	Number of Respondents					Average monthly earnings
	African	Asian	Coloured	White	Total	
Afrikaans	1283	132	19020	10277	30712	2752
English	448	5158	2569	6139	14314	4662
Sepedi	16110	0	24	3	16137	1835
Sesotho	17207	3	48	16	17274	1414
Setswana	18694	0	119	5	18818	1530
Xhosa	32336	4	94	46	32480	1502
Zulu	38583	16	45	10	38654	1594
Other than English	19842	258	170	326	20596	1739
Total/Overall	144503	5571	22089	16822	188985	2213

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.
HEAD	Dummy variable taking the value of 1 where the respondent is regarded as the head of the household.
ED1 ED2-4 ED5-7 ED8-10 EDTERT EDOTHER	Highest level of educational attainment. Linear regression splines were used to allow differing slopes across Standards 1-10, tertiary and other levels of educational attainment.
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.
AGE AGE ²	Age and Age ² to allow for a nonlinear effect. This would also capture Mincerian potential experience.
HOURS	Hours worked in the last 7 days.
Y1996 Y1997 Y1998	Time dummies to allow for different intercepts in each year.
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.
XHOSA ZULU SEPEDI SESOTHO SETSWANA AFRIKAANS OTHERL	Dummy variables taking the value of 1 where the language spoken at home is Isixhosa/Xhosa, Isizulu/Sizulu/Zulu, Sepedi/Northern Sotho, Setswana/Tswana, Afrikaans and a language other than these 7, respectively. English is the language of particular interest, ranking 8 th out of all possible responses. Hence, 8 categories were chosen. Base: English

Table 4
Participation
Double Hurdle & Heckman

	Official			Broad		
	Estimates	t statistic	Marginal effects	Estimates	t statistic	Marginal effects
C	-4.2822	-103.253	0.0000	-4.4936	-108.480	0.0000
ASIAN	-0.1956	-5.726	-0.0540	-0.3240	-9.320	-0.1048
COLOUR	0.1363	4.856	0.0428	0.0653	2.280	0.0235
WHITE	-0.1356	-4.666	-0.0384	-0.2342	-7.900	-0.0780
MALE	0.4512	63.778	0.1552	0.4180	59.050	0.1596
HEAD	0.4912	57.527	0.1705	0.3989	45.300	0.1520
ED1	0.0444	7.485	0.0135	0.0619	10.620	0.0223
ED24	-0.0307	-2.679	-0.0091	-0.0566	-5.010	-0.0198
ED57	-0.0573	-5.788	-0.0168	-0.0786	-8.070	-0.0274
ED810	0.2078	26.292	0.0668	0.2410	30.840	0.0899
EDTERT	0.1524	10.664	0.0481	0.0712	4.840	0.0257
EDOTHER	-0.7662	-10.250	-0.1603	-0.5559	-7.090	-0.1653
RURAL	-0.2973	-38.654	-0.0785	-0.2295	-29.960	-0.0765
MARRIED	0.0774	9.519	0.0238	0.0065	0.790	0.0023
XHOSA	-0.3939	-12.766	-0.0994	-0.2266	-7.200	-0.0756
ZULU	-0.1885	-6.149	-0.0522	-0.0298	-0.950	-0.0105
SEPEDI	-0.3169	-9.855	-0.0829	-0.2179	-6.660	-0.0729
SESOTHO	-0.1576	-4.987	-0.0442	-0.0555	-1.720	-0.0194
SETSWANA	-0.1835	-5.831	-0.0509	-0.0237	-0.740	-0.0084
AFRIKAANS	-0.0547	-3.009	-0.0160	-0.0585	-3.160	-0.0205
OTHERL	-0.1638	-5.319	-0.0459	-0.1036	-3.290	-0.0358
AGE	2.1732	156.485	0.4864	2.5077	183.550	0.7900
AGE2	-0.2708	-161.519	-0.0606	-0.3158	-189.610	-0.0995
Y1997	0.0025	0.302	0.0007	0.0200	2.400	0.0071
Y1998	0.1358	14.391	0.0426	0.1059	11.260	0.0385

Table 5
Employment
Double Hurdle Model

	Official			Broad		
	Estimates	t statistic	Marginal effects	Estimates	t statistic	Marginal effects
C	-0.9545	-12.561	0.1699	-1.4479	-22.268	0.0738
ASIAN	0.2160	3.714	0.0556	0.2601	5.049	0.0941
COLOUR	0.2963	6.543	0.0734	0.3235	8.300	0.1153
WHITE	0.6780	13.647	0.1378	0.6942	16.085	0.2212
MALE	0.1541	13.163	0.0408	0.2578	26.651	0.0933
HEAD	0.5856	43.792	0.1251	0.6152	55.822	0.2013
ED1	-0.0339	-3.358	-0.0097	-0.0097	-1.241	-0.0037
ED24	0.0394	2.080	0.0110	0.0215	1.461	0.0082
ED57	-0.0262	-1.624	-0.0075	-0.0178	-1.380	-0.0068
ED810	0.0836	6.524	0.0228	0.0904	8.458	0.0339
EDTERT	0.3614	15.365	0.0867	0.4428	21.023	0.1528
EDOTHER	-0.8625	-6.746	-0.3099	-1.2169	-11.311	-0.4390
RURAL	0.0436	3.460	0.0121	-0.1214	-11.736	-0.0470
XHOSA	-0.6238	-12.051	-0.2153	-0.7493	-16.598	-0.2914
ZULU	-0.4845	-9.445	-0.1618	-0.5784	-12.915	-0.2275
SEPEDI	-0.6094	-11.348	-0.2096	-0.6662	-14.246	-0.2609
SESOTHO	-0.4056	-7.700	-0.1325	-0.4776	-10.405	-0.1883
SETSWANA	-0.3750	-7.132	-0.1214	-0.5017	-10.981	-0.1978
AFRIKAANS	-0.0800	-2.269	-0.0234	-0.0897	-2.871	-0.0346
OTHERL	-0.4392	-8.459	-0.1449	-0.4547	-10.026	-0.1793
AGE	0.7488	26.601	0.7317	0.7394	31.074	0.4586
AGE2	-0.0650	-18.317	-0.0635	-0.0630	-21.022	-0.0391
Y1997	-0.0309	-2.181	-0.0089	-0.0312	-2.680	-0.0119
Y1998	-0.1504	-9.954	-0.0452	-0.0479	-3.776	-0.0184

Table 6
Employment
Heckman Model

	Official			Broad		
	Estimates	t statistic	Marginal effects	Estimates	t statistic	Marginal effects
C	-0.7763	-2.506	0.2188	-3.4204	-14.616	0.0003
ASIAN	0.2206	3.762	0.0541	0.1738	3.307	0.0692
COLOUR	0.2924	6.393	0.0692	0.3439	8.798	0.1363
WHITE	0.6813	13.636	0.1309	0.6343	14.495	0.2440
MALE	0.1414	5.822	0.0360	0.3747	22.662	0.1482
HEAD	0.5729	22.751	0.1169	0.7135	45.155	0.2711
ED1	-0.0351	-3.450	-0.0097	0.0069	0.868	0.0027
ED24	0.0402	2.143	0.0107	0.0066	0.451	0.0026
ED57	-0.0245	-1.498	-0.0067	-0.0413	-3.131	-0.0163
ED810	0.0776	4.733	0.0204	0.1603	11.960	0.0638
EDTERT	0.3591	15.097	0.0822	0.4441	20.978	0.1747
EDOTHER	-0.8459	-6.467	-0.2975	-1.3329	-12.247	-0.3802
RURAL	0.0520	2.740	0.0138	-0.1851	-14.608	-0.0722
XHOSA	-0.6130	-11.177	-0.2058	-0.8128	-17.735	-0.2774
ZULU	-0.4793	-9.217	-0.1553	-0.5881	-13.100	-0.2132
SEPEDI	-0.6010	-10.820	-0.2012	-0.7278	-15.357	-0.2545
SESOTHO	-0.4015	-7.560	-0.1271	-0.4943	-10.739	-0.1830
SETSWANA	-0.3700	-6.954	-0.1159	-0.5101	-11.140	-0.1882
AFRIKAANS	-0.0787	-2.227	-0.0221	-0.1077	-3.431	-0.0423
OTHERL	-0.4350	-8.308	-0.1391	-0.4845	-10.633	-0.1798
AGE	0.6766	5.407	0.5476	1.6064	15.827	0.7293
AGE2	-0.0559	-3.565	-0.0452	-0.1732	-13.436	-0.0786
Y1997	-0.0308	-2.175	-0.0085	-0.0256	-2.197	-0.0101
Y1998	-0.1538	-9.514	-0.0446	-0.0181	-1.382	-0.0072
λ^P	-0.0470	-0.594	-0.0130	0.5407	8.733	0.2106

Table 7
Income
Double Hurdle Model

	Official		Broad	
	Estimates	t statistic	Estimates	t statistic
C	4.3683	26.500	4.3549	28.010
ASIAN	0.0143	0.440	0.0089	0.270
COLOUR	0.0497	1.860	0.0457	1.710
WHITE	0.4157	14.710	0.4200	15.070
MALE	0.3919	28.540	0.3775	31.730
HEAD	0.1065	6.740	0.0994	6.830
ED1	0.0709	10.350	0.0694	10.240
ED24	-0.0138	-1.070	-0.0136	-1.060
ED57	0.0243	2.150	0.0220	1.940
ED810	0.0873	8.550	0.0883	8.720
EDTERT	0.1720	14.670	0.1687	14.350
EDOTHER	-1.1025	-18.510	-1.0809	-18.230
RURAL	-0.4181	-35.440	-0.3944	-39.820
MARRIED	0.0939	11.460	0.0865	10.680
XHOSA	-0.2411	-7.750	-0.2244	-7.220
ZULU	-0.1407	-4.790	-0.1283	-4.340
SEPEDI	-0.1291	-4.000	-0.1247	-3.880
SESOTHO	-0.4065	-13.640	-0.3993	-13.360
SETSWANA	-0.2337	-7.870	-0.2167	-7.230
AFRIKAANS	-0.1997	-12.760	-0.2022	-12.910
OTHERL	-0.1308	-4.470	-0.1321	-4.540
AGE	0.8597	13.110	0.8939	13.620
AGE2	-0.0922	-11.200	-0.0965	-11.680
HOURS	0.0013	5.470	0.0013	5.490
INFML	-0.4351	-49.740	-0.4339	-49.550
Y1997	0.1118	12.430	0.1129	12.530
Y1998	0.2082	18.080	0.1938	18.380
λ^P	0.1762	3.750	0.1739	4.180
λ^E	-0.2402	-4.500	-0.1585	-4.170

Table 8
Income
Heckman Model

	Official		Broad	
	Estimates	t statistic	Estimates	t statistic
C	4.9139	65.690	4.9670	66.030
ASIAN	0.0355	1.110	0.0317	0.990
COLOUR	0.0445	1.670	0.0396	1.480
WHITE	0.4376	15.810	0.4359	15.890
MALE	0.3521	40.660	0.3430	37.340
HEAD	0.0794	5.710	0.0707	5.170
ED1	0.0653	9.760	0.0642	9.640
ED24	-0.0093	-0.720	-0.0088	-0.680
ED57	0.0291	2.590	0.0288	2.570
ED810	0.0703	7.700	0.0682	7.460
EDTERT	0.1720	14.660	0.1697	14.480
EDOTHER	-1.0669	-18.150	-1.0519	-17.820
RURAL	-0.3883	-44.510	-0.3756	-41.420
MARRIED	0.0887	10.970	0.0881	10.900
XHOSA	-0.2223	-7.260	-0.2079	-6.700
ZULU	-0.1373	-4.680	-0.1272	-4.310
SEPEDI	-0.1172	-3.650	-0.1089	-3.400
SESOTHO	-0.4033	-13.540	-0.3963	-13.260
SETSWANA	-0.2272	-7.660	-0.2160	-7.210
AFRIKAANS	-0.1972	-12.610	-0.1983	-12.680
OTHERL	-0.1278	-4.370	-0.1250	-4.300
AGE	0.6345	25.030	0.6277	25.880
AGE2	-0.0633	-22.770	-0.0628	-23.420
HOURS	0.0013	5.420	0.0013	5.470
INFML	-0.4345	-49.670	-0.4333	-49.470
Y1997	0.1110	12.340	0.1114	12.390
Y1998	0.1918	17.970	0.1858	17.950
λ^E	-0.1668	-3.340	-0.1591	-4.260