

**DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH**

**THE HUMAN CAPITAL THEORY AND THE SORTING
HYPOTHESIS: EXPLAINING LABOUR MARKET RETURNS
TO EDUCATION IN SOUTH AFRICA**

by

Hendrik van Broekhuizen

Assignment presented in partial fulfillment of the requirements for the degree of Bachelors of Commerce with Honours at the University of Stellenbosch.

Supervisor: Dr. P de Villiers

March 2009

Table of Contents

List of tables.....	i
List of Figures.....	i
1. INTRODUCTION.....	1
2. HUMAN CAPITAL THEORY AND THE SORTING HYPOTHESIS.....	4
2.1 The Concept of “Human Capital”.....	4
2.2 The Human Capital Theory (HCT) Explained.....	5
2.3 The Sorting Hypothesis (SH) Explained.....	5
2.4 Screening and Signalling Explained.....	7
3. THE RELEVANCE OF THE HCT-SH DEBATE AND CONSEQUENCES OF THE RESPECTIVE THEORIES.....	9
4. TESTING THE HCT AND THE SH.....	12
4.1 Issues Encountered in Testing the HCT and the SH.....	12
4.2 Previous Tests of the SH.....	13
5. TESTING THE SORTING HYPOTHESIS FOR SOUTH AFRICA.....	16
5.1 Empirical Methodology and Data.....	16
5.2 The Empirical Model.....	18
5.3 Empirical Specification and Results.....	19
6. CONCLUSION.....	30
REFERENCES.....	31

List of Tables

Table 1 - Hypotheses for testing the HCT and SH.....	16
Table 2 - Multinomial Logits for Black Individuals.....	21
Table 3 - Multinomial Logits for Coloured & Indian Individuals.....	22
Table 4 - Multinomial Logits for White Individuals.....	23
Table 5 - MNL Corrected Mincerian earnings functions (log of wages) for Black Individuals.....	24
Table 6 - MNL Corrected Mincerian earnings functions (log of wages) for Coloured Individuals.....	25
Table 7 - MNL Corrected Mincerian earnings functions (log of wages) for White Individuals.....	26
Table 8 - Summarised findings from empirical tests of the SH in the South African labour market.....	29

List of Figures

Figure 1 - Returns to Education for Black Individuals.....	27
Figure 2 - Returns to Education for Coloured Individuals.....	27
Figure 3 - Returns to Education for White Individuals.....	27

1. INTRODUCTION

The positive relationship between formal education, productivity and labour market earnings is one of the most frequently studied empirical findings in economics (Chatterji, Seaman and Singell, 2003:191). Studies investigating the rates of return (RORs) to education are prolific and while they may differ in many regards, they share commonality in finding that, on average, higher levels of education are associated with higher labour market earnings. Similarly, it is found that education generally shares a positive association with the probability of employment with the more educated not only more likely to procure employment than the less educated, but also more likely to procure better employment (Bhorat & McCord, 2003:135).

The reasons why education shares a positive association with both earnings capacity and the probability of employment are tantamount and relate to the positive correlation between education and productivity (Chatterji, Seaman and Singell, 2003:191). Employers, desiring high levels of productive output, are not only more inclined to employ high-productivity workers, but also to remunerate high-productivity labour better than its less productive counterpart. Consequently, from the frequently observed positive correlation between earnings and education it may be inferred that education and productivity share a similarly strong and positive correlation (Psacharopoulos, 1995:12). However, while there is a general consensus regarding the existence of such a positive correlation, the causal relationship between education and productivity is an issue of considerable and often controversial debate (Jaag, 2006:7). Specifically, it is uncertain whether causality runs from education to productivity or from innate productivity to education or if, to some extent, education and productivity share a mutually reinforcing relationship with causality running in both directions.

The two dominant theories that venture to establish the nature of causality between education and productivity and thereby provide a theoretical justification for the correlation between education and earnings are the Human Capital Theory (HCT) and the Sorting Hypothesis (SH) (Haspel, 1978:280).¹ While the HCT contends that education influences earnings directly by raising the productivity of workers through the augmentation of their innate capacities and installation of relevant skills, the SH extends the HCT by allowing for

¹ The Sorting Hypothesis encompasses both the Screening- and Signalling Hypothesis. A discussion of these two components of the SH is presented in Section 2.4 below.

education to influence earnings indirectly through acting as a carrier of inherent productivity-related information. Both of these theories are utilised by firms in their search for employees with superior levels of productivity and both entail that job-seekers acquire the highest attainable level of education in an attempt to raise their labour market earnings potential and increase the likelihood of procuring employment (Koch and Ntege, 2006a:2).

The fact that both the HCT and SH are consistent with observing a positive correlation between education and earnings renders the debate over their respective validities fairly redundant at the level of the individual (Brown and Sessions, 1998:586). Since both theories contend that education raises lifetime earnings and the probability of employment, individuals may still find it “profitable” to invest in education irrespective of which theory dominates. However, whether education represents a “profitable” investment for society is less clear. If education is found to be primarily productive, as per the HCT, then social spending on education is justified since greater investment in education leads to greater aggregate productivity for society. Subsequently, under the HCT it is expected that the private and social RORs to education will be proportionate.² If, however, education’s function is predominantly informational, investment in education is not expected to raise aggregate productivity, such that its private and social RORs will be disproportionate. These two theories thus have important implications, not only for whether education may be a good investment or not, but, more crucially, for policy issues related to who should incur the costs of education.

Very little work has thus far been done to empirically establish the extent to which the HCT and SH, respectively, prevail in the South African labour market.³ Perhaps this is not surprising since, in the case of South Africa, the already arduous task of attempting to distinguish between the HCT and SH components of the returns to education is further complicated by differentials in the quality of, and the labour market returns to, education across race groups (van der Berg, 2005:1). Therefore, Brown and Sessions’s (1998:398) assertion that the prevalence of sorting in a labour market depends on culture and institutional structure implies that, in a country still characterised by perverse and pervasive racial inequalities, sorting, if indeed prevalent, is likely to manifest to a varying extent across race

² The private and social rates of return to education under consideration throughout this paper are comprehensively defined in Section 3 below.

³ See Koch and Ntege (2006a, 2006b, and 2008).

groups.⁴ Consequently, when testing the empirical validity of sorting in South Africa, it would be prudent to allow for the analysis of both the remuneration-returns to education and selection into employment to be conducted separately for different race groups.

The aim of this paper is to provide a comprehensive overview of the HCT and SH and to empirically test the prevalence of sorting by education in the South African labour market. The empirical analysis departs from former studies of the SH by disaggregating multiple tests of sorting by race group and sector of employment. Utilising multinomial sample selection correction regression analysis, the Wolpin (1977), Psacharopoulos (1979), and Kroch and Sjoblom (1994) methods are used in conjunction to test sorting for Blacks, Coloureds, and Whites in the public-, private, and self-employed sectors of South Africa. These established methods remain the most frequently utilised methodologies to test for sorting and provide testable hypotheses that serve as indicators regarding the extent to which education performs its productivity-augmenting and innate productivity-reflecting functions in the labour market. The empirical results indicate that education performs both a productivity-augmenting and innate productivity-reflecting function in the South African labour market. Furthermore, sorting in South African is found to have a strong racial dimension, mattering most for Whites in the private sector, Blacks in the public sector, and least overall for the Coloured race group.

The paper proceeds as follows: Section 2 provides an overview of the concept of human capital, the Human Capital Theory, the Sorting Hypothesis, and the Screening and Signalling Hypotheses. Section 3 considers the relevance of the HCT-SH debate and the implications of either theory. Section 4 discusses some of the issues pertaining to empirically testing the SH and provides a brief literature review of some of the empirical methods that have been used to test its relevance. Section 5 presents the empirical methodology, model, estimation specification, and results for the empirical analysis of sorting in the South African labour market and Section 6 concludes.

⁴ The concept of “sorting” is used throughout this paper to refer to any act or manifestation whereby the innate-productivity reflecting function of formal education serves to sort or rank workers according to their desirable but unobserved characteristics and capacities to such an extent that some component of the return to education is effectively a return to the information implicitly contained in the level of educational attainment. It follows that a test for sorting is synonymous with a test of the HCT and SH.

2. HUMAN CAPITAL THEORY AND THE SORTING HYPOTHESIS

2.1 The Concept of “Human Capital”

The concept of human capital and its relevance with regard to education and labour market productivity was first pioneered in the early 1960’s by T.W. Schultz and later extended by Gary S. Becker and Jacob Mincer (Becker, 1992b:43). In the broadest sense, human capital refers to the sum of all physical, mental, and psychological personal experiences, characteristics and capacities that have bearing upon an individual’s potential and realized labour market productivity. However, there remains no single encompassing and consistent definition of human capital nor any explicit consensus of what it should consist of. Although it was initially the subject of much controversy, the term “human capital” is now generally accepted and frequently used by economists and other social scientists (Becker, 1992a:85).

Every individual possesses a stock of human capital that is comprised of a unique blend of many constituent components. Consequently, the worth of any given stock of human capital will depend critically on the nature of the individual’s occupation (Wolpin, 1977:950). As the requirements of a job become more aligned with the type of human capital that an individual possesses, the value of that human capital will rise and the individual’s realized labour market productivity will converge on his/her potential labour market productivity. The implication is that individuals with certain sets of skills and attributes may be highly productive in occupations where those skills and attributes are valued, but considerably less productive in those jobs where they are not.

Due to homonymy, it is perhaps necessary to clarify that the concept of human capital is not exclusive to the HCT. Human capital simply encompasses all the factors within the individual that affect productivity and, as such, is consistent with both the HCT and SH.⁵ However, the two theories differ in their assertions regarding whether human capital is mutable, and if so, to what extent it may be augmented through formal education.

⁵ “Human capital” is often used to refer to the sum of a group of individuals’ stocks of human capital or as a generic collective for a work force. While such interpretations are simply derivatives of the one presented here, the use of “human capital” in this paper, unless stated otherwise, refers to the human capital of an individual.

2.2 The Human Capital Theory Explained

The HCT postulates that education's primary function is to augment and expand individuals' stocks of capacities, skills, and personality traits over the duration of the education process. More specifically, the HCT contends that education, through the process of learning, enhances those factors that amount to, lead to, or are associated with, ultimate labour market productivity. Thus, education does not simply impart irrelevant skills and knowledge, but acts as both an augments of innate human capital and a creator of new human capital such that an investment in education is an investment in human capital. According to this theory, individuals will invest in their education until the gain in their marginal labour market productivity equals the marginal cost incurred in acquiring additional education (Kim and Sakamoto, 2005:4). The main contention of the HCT is therefore that causality runs from education to productivity to earnings and, subsequently, that the labour market returns to education are a consequence of its productivity-augmenting function.

2.3 The Sorting Hypothesis Explained

The SH was pioneered by economists such as Berg (1970), Arrow (1973), Spence (1973), Taubman and Wales (1973), Stiglitz (1975) and Wolpin (1977) and contends that the primary function of education is to reflect to employers in the labour market the innate productivity-related characteristics of workers. The SH is often mistakenly interpreted to be the antithesis of the HCT. Such interpretations possibly stem from the fact that much of the literature on the extreme versions of the SH inadvertently creates the impression that the HCT and SH are always and everywhere rival and mutually exclusive hypotheses (Weiss, 1995:134).

The most stringent interpretation of the SH, called the Strong Screening Hypothesis (SSH), does indeed claim that productivity is entirely innate and therefore immutable (Layard and Psacharopoulos, 1974:986). Consequently, education is presumed to have no other function than to provide individuals with credentials that reflect their innate productivities to employers, where higher levels of educational attainment are associated with higher innate ability.⁶ Understandably, this interpretation is viewed by many as an attack on the validity of the HCT and is therefore met by considerable opposition from the proponents of the HCT. However, although it is often presented as the standard version of the SH, such an agnostic

⁶ Here, "educational attainment" encompasses the quantity, quality and content of an individual's education as well as the individual's educational achievement.

view is an extreme interpretation of sorting that remains largely unsubstantiated empirically (Brown and Session, 1998:587).

In contention to the view that the HCT and the SH are inherently irreconcilable, Weiss (1995:134) asserts that the only difference between the SH and HCT is that sorting allows “...for some productivity differences that firms do not observe to be correlated with the costs or benefits of schooling.” In essence, the SH thus asserts that there is some part of an individual’s human capital that is inherent and immutable. Consequently, while the SH shares commonality with the HCT in allowing for the expansion of productivity through learning, its main assertion is that there are a multitude of innate productivity-related attributes, which employers cannot readily observe at the point of employment, that are not augmented by, but are correlated with the education process. Here, the “point of employment” refers to the time at which a job-applicant becomes employed, the initial period of his/her employment, and the managerial decision regarding the new employee’s salary-offer.

Attributes such as a lower propensity to quit, scrupulousness in successfully completing tasks, strong self-motivation and drive, the capacity to grasp and internalise complex information and concepts, emotional maturity, and the ability to work well under pressure have important bearing on productivity and are therefore of considerable interest to employers. However, such characteristics are difficult to observe directly without long periods of on-the-job monitoring and are therefore largely shrouded from employers at the point of employment. The SH therefore maintains that education is correlated to such an extent with these indispensable characteristics that an individual’s educational attainment may reveal to employers the extent to which they are present in that individual. (Weiss, 1995)

It follows from the above that the feasibility of the SH critically depends on educational attainment’s ability to effectively rank individuals according to the presence of certain characteristics and, hence, their innate productivities. However, the SH’s conjecture that high levels of educational attainment are a signal of high levels of innate productivity can only hold if the cost of acquiring education (tuition fees and other physical expenses as well as the psychological and intangible opportunity costs associated with attaining education) is inversely proportional to an individual’s natural ability. If this were not the case, all individuals would be able to invest in education to the same extent, thereby making it

impossible to distinguish between their productive capabilities on the basis of their educational attainment (Spence, 1973:358).

The education process commonly involves a series of competency based tasks and tests of natural abilities that are indented to separate individuals according to their levels of performance and, consequently, ability. It is found that individuals with higher abilities not only find it less costly to perform these tasks and tests and thereby progress through the education system, but that they are subsequently also likely to acquire levels of education in excess of their less able counterparts (Weis, 1995:137). Educational attainment thus satisfies the criteria for an effective sorting device. Moreover, the high level of legitimacy that education possesses as a factor in the framework of statistical discrimination, when used to determine an individual's position within certain probabilistic productivity distributions, promotes its use as a sorting device (Cohn and Geske, 1990:60).

The SH thus holds that employers can overcome the imperfect information they possess regarding individuals' levels of productivity by inferring it from their levels of educational attainment. Education subsequently acts as a filter that enables employers to sort and rank individuals according to these inferred levels of productivity. Causality, in addition to running from education to productivity, thus also runs from innate productivity to education. The SH therefore allows for the labour market returns to education to stem partially from its productivity-augmenting function and partially from its innate productivity-reflecting function.

2.4 Screening and Signalling Explained

The SH encompasses both the Signalling Hypothesis and the Screening Hypothesis. The primary difference between these two theories is game-theoretical in nature: Under signalling, workers (who are informed about their own productivity) acquire education with the purpose of "signalling" their innate abilities to prospective employers. Under screening, employers (who are uninformed about the productivity of workers) may demand minimum levels of education from applicants or simply look to their given levels of educational attainment in order to "screen" prospective employees and infer their innate abilities. Thus, while in both theories the aim of education is to inform employers about employee productivity, it is the employees who are presumed to act first to disclose this information under the Signalling Hypothesis, whereas employers act first in attempting to infer productivity under the Screening Hypothesis. (Stiglitz and Weiss, 1990:1)

In the context of education, signalling and screening share a special kind of dichotomy. Screening on the basis of an individual's educational attainment can only be effective if educational attainment has some signalling power. Similarly, individuals, hoping to signal their innate abilities to potential employers, can only do so if employers, to some extent, consider educational attainment as a device that screening can be based on. As a result, the convention that will be used throughout this paper is to abstract from game-theoretical considerations regarding who is responsible for revealing relevant information with respect to individuals' innate productivities in the labour market, and to refer to signalling, screening and sorting interchangeably.⁷

Using the convention above, two versions of the SH can be distinguished. These are derived from Psacharopoulos (1979)'s Strong Screening Hypothesis and Weak Screening Hypothesis and are referred to here, respectively, as the Strong Sorting Hypothesis (SSH) and the Weak Sorting Hypothesis (WSH). Under the SSH, productivity is held to be entirely innate and is therefore immutable by education. Education is presumed, however, to capture all innate productivity-related information perfectly such that educational attainment is a wholly sufficient signal or screen. The implication is that employers are fully informed about individuals' actual labour market productivities based on their levels of educational attainment so that no new information about the productivity of their workers can be revealed through on-the-job observation (Koch and Ntege, 2008:15). The SSH thus predicts that employers will set the wages of their employees perfectly, such that a sorted employee's earnings-returns to education will remain constant over the entire span of his/her employment.

The WSH relaxes the stringent assumptions of the SSH and allows for educational attainment to be an imperfect signal or screen. The WSH is thus consistent with the assumptions of the SH discussed in Section 2.3 above and purports that education does perform some sorting function, but that it also retains some productivity-augmenting role. Consequently, employers may set initial wages proportional to individuals' inferred innate productivities, but will adjust those wages as they become more aware of employees' realized productivities through on-the-job observation. In essence, the return to education thus "transfers" to experience as employers replace their dependency on the signal of education by gauging

⁷ It is important to note that signalling and screening do have distinct theoretical implications for labour market equilibria. However, the purpose of this paper is to deal with the empirics of sorting and therefore abstracts from such theoretical considerations. For a comprehensive discussion on the theoretical similarities and differences between the Signalling Hypothesis and the Screening Hypothesis see Stiglitz and Weiss (1990).

worker productivity from experience. This implies that the earnings-returns to education for a sorted employee should decline with his/her experience or tenure. (Magoula and Psacharopoulos, 1999:1594)

It is critical to emphasise that the HCT renders the same prediction as the SSH with respect to the relationship between the returns to education and experience or tenure. If education augments productivity, as per the HCT, then one would not expect the wage-returns to education to fall as an employee gains experience in a given job.⁸ This prediction is consistent with the SSH and therefore presents a conundrum for any attempts to empirically distinguish between the two theories. However, because the SSH is simply a more stringent version of the WSH, evidence that would seem to confirm the SSH must first be corroborated by evidence that confirms the WSH before one could conclude that support is found for the SSH. If this is not the case, constant returns to educational attainment over time would provide more support for the HCT than for the SH.

3. THE RELEVANCE OF THE HCT-SH DEBATE AND CONSEQUENCES OF THE RESPECTIVE THEORIES

While the SH may be regarded as an extension of the HCT, the two theories have vastly different implications for education choices, education policy, and social welfare (Brown and Sessions, 1998:586). The main relevance of the HCT-SH debate with respect to education policy centres around the socially desirable level of public spending on education. It is important to clarify at this point that the concepts of “returns to education” and “rates of return to education” that are used below are very narrowly and specifically defined to focus exclusively on the monetary and/or productive benefits of education. The private ROR to education is defined as the marginal increment in an individual’s wages (in percentage) that results from acquiring more education. As such, the private ROR to education considered here abstracts from all other direct and indirect non-wage returns to education. The private ROR to having a Masters degree, for example, is the percentage wage-premium that an individual receives over the wage that he/she would have received if he/she had no schooling, *ceteris paribus*.

⁸ In fact, since the HCT presumes that the wage-returns to education are, in essence, the returns to that portion of productivity that is “created” by education, the returns to education could only fall over time if an employee’s productivity fell over time.

The social ROR to education is defined as the marginal increment in a country's productive output (percentage rise in GDP) that results from all individuals in that country acquiring more education. This definition abstracts from all indirect social benefits of education such as individuals' increased capacities to pay taxes and focuses entirely on increases in aggregate economic productivity that are direct consequences of education making individuals more productive.

If education is entirely a productivity-enhancing process, as the HCT suggests, investment in education would generate private returns in the form of increased personal labour market earnings and social returns in the form of increased aggregate productivity in the economy. Such a result ensues because the returns to education are effectively returns to education-augmented productivity. Thus, if the HCT holds, one would expect the private and social RORs to education to be proportionate. (Quiggin, 1999:134)

In contrast to assertions of the HCT, it is often argued that, if education's function is informational, the private and social RORs to education will diverge, with the social return being substantially lower than the private return (Weiss, 1995:136). This is because, under the SH, the private return to education accrues partially to education's productivity-augmenting role, but also partially to its innate-productivity reflecting role. However, from the definitions given above it is clear that the social ROR to education is generated entirely by education's productivity-augmenting function. The implication is that, under the SH, the private returns to education would be raised above the social returns by that portion of the private return that results from education's information revealing function.

It is clear from the arguments above that investment in education may be privately beneficial, regardless of whether the HCT or the SH dominates in the labour market, but that the two theories have very different implications for the social returns to education. Under the former theory the private and social returns to education are proportional, but under the latter, private individuals stand to benefit more from their own education than society does. The standard argument for policy in this respect is therefore that social spending on education remains justified as long as education is found to be socially productive. That is: as long as the HCT holds. By extension, the greater the empirical validity of the SH, the greater the proportion of education costs that should, arguably, be borne privately. (Weis, 1995)

In contrast to the former argument, Weiss (1995:134) reiterates Wolpin (1977)'s contention that the prevalence of sorting in a labour market could give rise to social RORs to education

that equal or even exceed the private RORs, such that public spending on education would be justified even in the presence of sorting. To explain this assertion, consider again the discussion on the concept of human capital presented in Section 2.1 above. The critical issue here is that individuals are presumed to be more productive in jobs that are well-suited to the skills they possess than in those jobs that are not. Therefore, if it holds that individuals' realized labour market productivities converge on their potential labour market productivities as their occupations become more aligned with their type of human capital, then there is some productive benefit to be derived from correctly allocating individuals to suitable occupations, over-and-above that of education's direct productivity-enhancing function. Therefore, provided that there is no other mechanism that would be more efficient than sorting in appropriately allocating individuals to suitable occupations, the innate productivity-reflecting function of education may well raise the productivity of individuals and, therefore, raise the social returns to education.

Even if Wolpin (1977)'s argument holds true, the SH's potential implications for education choice may again erode its social returns and render the former argument somewhat redundant. Under the theory of screening, an expansion of the pool of tertiary educated individuals will dilute the significance of tertiary credentials such that employers will revise their hiring criteria upwards (Blaug, 1984:21). The implication is that individuals may have to acquire excessive levels of education in order to meet the new hiring standards. Consequently, sorting is likely to disrupt education choices, distort the social and private returns to education and lead to Pareto inefficient outcomes in the labour market.

Finally, the HCT-SH debate has particular relevance in South Africa since, depending on which theory is found to be more relevant, education can play a significant role in enforcing inequalities (van der Berg, 2005:1). If the HCT holds, heterogeneity in the quality of education will cause heterogeneity in the quality of its human capital output. Under this scheme, education may thus actually exacerbate inequalities in the labour market. By contrast, the SH contends that education performs less of a productivity-enhancing role and may therefore, at first glance, seem to have more benign implications labour market inequalities. However, under the SH, the same heterogeneous educational quality implies that the ability filtering function of education produces signals of heterogeneous quality. Consequently, while an individual attending an educational institution of poor quality may be employable under the HCT by at least acquiring new human capital to some extent, the same

individual may be precluded from employment under the SH if employers lack confidence in quality of the signal inherent in his/her educational attainment.

4. TESTING THE HCT AND THE SH

4.1 Issues Encountered in Testing the HCT and the SH

Both the HCT and the SH ascribe the positive association between education and labour market earnings to the positive relationship between education and productivity and, as such, are consistent with observing positive earnings-returns to education. These predictive similarities imply that it is not only difficult to distinguish between the two theories empirically, but that it is even more arduous to decompose the returns to education into the portion that accrues to education's productivity-augmenting function and the portion that accrues to its innate productivity-reflecting function (Koch and Ntege, 2006b:143).

An intuitively appealing method for performing the empirical disaggregation between the HCT and the SH would be to capture individuals' on-the-job productivities before they undertake education and then again thereafter. The extent to which these *ex ante* and *ex post* productivities diverge would then, arguably, yield an indication of the extent to which education enhances labour market productivity. Furthermore, in the knowledge of the *ex ante* and *ex post* productivities, a comparison of the *ex ante* and *ex post* labour market earnings would indicate what portion of the returns to education accrues to education-augmented productivity and what portion accrues to education-reflected information.

Clearly, the method described above is not practically feasible since the temporal nature of the education-to-employment relationship implies that *ex ante* productivities and earnings are hypothetical and, therefore, unobservable. To overcome this issue, many researchers have tried to use ability measures such as IQ scores as proxies for *ex ante* productivities in an attempt to control for that portion of the returns to education that accrues to innate ability. However, the inclusion of such proxy measures are seldom found to significantly impact the returns to education in earnings regression studies (Weis, 1995:235). In fact, the extent of sorting in the labour market could only be established through this method if it holds that employers possess poorer information regarding employees' innate abilities than the researchers (economists, sociologists, etc.) utilising the method do (Wolpin, 1977:950). If

this were not the case, employers would have little incentive to sort employees on the basis of educational attainment since they would be aware, to some extent, of their innate abilities. The increased use of psychometric- and other ability-based tests at the point of employment imply that it is unlikely that employers would have poorer knowledge of individuals' abilities than researchers who investigate sorting do.

Tests of the SH and HCT are further complicated in countries where labour market and/or educational inequalities distort the earnings-returns to education. Such is the case in South Africa where labour market discrimination along racial and gender lines remains prevalent (Burger and Jafta, 2006:31). Furthermore, van der Berg (2005:1) finds that the South African education system is characterised by differential access to quality education across race groups. The joint implication of these two phenomena is that the coefficients of education variables, found in regression analyses, may be biased estimates of the actual labour market wage-returns to education in South Africa. While it may not be practically possible to fully control for these aspects when conducting tests of the SH for South Africa, it would be prudent to allow for different race groups to have different data generating processes with respect to the estimation and comparison of Mincerian earnings functions.

4.2 Previous Tests of the SH

In practice, the most common method for testing the relevance of the HCT and the SH involves separating individuals into groups in which sorting would be beneficial or likely to occur and those in which it would not be beneficial or would be unlikely, and then to compare the RORs to education for the two groups (Kroch & Sjoblom, 1994:157). Following this methodology, the latter group would constitute the "unsorted" sample for which the returns to education accrue entirely to its productivity-enhancing function. The former group would constitute the "sorted" sample for which the returns to education should also partially reflect its information unveiling function. Consequently, by comparing the RORs to education for these two groups, one could derive estimates of the extent to which the SH and HCT, respectively, apply in a labour market. Two obvious deficiencies present themselves in this method. The first involves the degree of subjectivity that is needed to identify the "sorted" and "unsorted" groups for analysis, and the second relates to the fact that, since the "sorted" sample serves as the control group against which the prevalence of sorting is tested for the other group, the method requires that some group in the labour market

must be excluded from the test for sorting. Despite these issues, this methodology and the variations thereof remain the most commonly used tests of the SH and HCT.

Psacharopoulos (1979) assigns the competitive (private) sector of the economy as the “unsorted” control group and the uncompetitive (public) sector as the “sorted” group (Arabsheibani and Rees, 1998:189). The justification for this assignment rests on the assumption that wages in the public sector “...are more likely to be determined in a bureaucratic and centralised manner and thus tend to be more rigidly fixed to years of schooling” whereas wages in the private sector are more likely to be closely linked to an employee’s marginal productivity (Clark, 2000:6). By comparing the RORs for the two sectors, Psacharopoulos (1979) is thus able to test for the validity of the WSH. Furthermore, by examining the relationship between education and tenure, an indication of the extent to which the SSH holds is derived.

One of the major flaws of the Psacharopoulos (1979) method is that it effectively entails that the prevalence of sorting in the private sector cannot be tested. The Wolpin (1977) method therefore departs from the Psacharopoulos (1979) method by arguing that sorting is likely to occur in both the private and the public sectors of the economy, but not in the sector for the self-employed. The rationale for this assumption is that, since the self-employed are not faced with the agent-principal goal deviation problem like their employed counterparts, they would presumably only invest in education for its productivity augmenting function (Koch and Ntege, 2006a:11). Consequently, the rate of return to education for the self-employed should accrue entirely to its productivity-enhancing role. Therefore, RORs to education in the employed sectors in excess of those in the self-employed sector would provide evidence in support of the SH (Koch and Ntege, 2006a:12).

Empirical applications of the Psacharopoulos (1979) and Wolpin (1977) methods for various countries have tended to provide evidence against the SSH and render equivocal results for the WSH and the HCT.⁹ The only attempts to test the prevalence of sorting in the South African labour market have been by Koch and Ntege (2006a, 2006b and 2008). Employing both the Psacharopoulos (1979) and Wolpin (1977) methodologies and using data from Statistics South Africa’s Labour Force Surveys for the whole South African population, the authors find some broad support for the WSH in all three of their studies and conclude that

⁹ See Arabsheibani and Rees (1998), Brown and Sessions (1998, 1999 and 2005), Clark (2000), and Castagnetti, Chelli and Rosti (2005).

education performs both a sorting and a productivity-enhancing role in South Africa. However, none of these three studies make allowance for the fact that the prevalence of sorting may differ according to race and instead are conducted for the South African population as a whole.

Numerous additional tests of the SH have been developed, either as extensions of the Psacharopoulos (1979) and Wolpin (1977) methodologies or in response to their respective deficiencies. Riley (1979) contrasts the returns to education between occupations in which sorting is more likely to occur and those occupations where it is not and finds support for the SH relative to the HCT. Liu and Wong (1982) estimate the RORs to education credentials and construct mid-to-early career earnings ratios in order to test for the SSH. Ashenfelter and Krueger (1992) utilise a study of identical twins in an attempt to isolate the relationship between education and earnings when ability and family background are held constant and find evidence in support of the HCT. Groot and Oosterbeek (1994) divide educational attainment into effective years, repeated years, skipped years, inefficient routing years and dropout years. By testing the coefficients of these variables against the predictions of the HCT and the SH they find strong support for the HCT relative to the SH. While this approach is certainly very appealing it relies on very specific and tailored data, the kind of which is seldom encountered.

Lastly, Kroch and Sjoblom (1994) argue that sorting, if relevant, should manifest with regard to an individual's ranking in the educational attainment distribution for his/her age-cohort and not in terms of absolute levels of educational attainment. By estimating earnings regressions with alternate specifications, the authors test whether an individual's absolute educational attainment (in years) generates more significant earnings-returns than his/her rank in the relevant educational attainment distribution. Consequently, by comparing the coefficients on the two alternate measures of education to the predictions of the HCT and SH, the authors are able to attest to the respective magnitudes of education's sorting and productivity-augmenting roles.

5. TESTING THE SORTING HYPOTHESIS FOR SOUTH AFRICA

5.1 Empirical Methodology and Data

Attempts to determine the empirical validity of the HCT and SH generally rely on the results from a single test of sorting to reveal which of the theories dominates in a labour market. However, tests of the SH are all fallible to some degree and their results can at best be viewed as indicators of the presence of sorting. Consequently, the findings from any study examining the empirical validity of the HCT and SH would be more robust if it were based on more than one method for testing the SH. If different methods yield similar results, those results may more readily be interpreted as confirmation of whether education augments productivity, reflects innate productivity, or performs a combination of these functions.

In the analysis that follows, the Wolpin (1977), Psacharopoulos (1979) and Kroch and Sjoblom (1994) methods are used in conjunction to test for sorting in the South African labour market. Each of these methods provides at least one testable hypothesis for establishing whether the labour market is indeed characterised by some degree of sorting. Table 1 below presents the respective hypotheses, the theory that would be supported if a given hypothesis is found to hold, and the mnemonic devices that will be used to refer to the corresponding hypotheses in the discussion that follows.

Table 1 - Hypotheses for testing the HCT and SH

Method	Mnemonic	Hypothesis	Theory supported
Wolpin (1977)	<i>W(I)</i>	RORs to education are higher for the employed than for the self-employed	WSH
Psachharopoulos (1979)	<i>P(I)</i>	RORs to education are higher for the employed in the public sector than in the private sector	WSH
Psachharopoulos (1979)	<i>P(II)</i>	RORs to education do not change with experience/tenure	SSH
Kroch and Sjoblom (1994)	<i>SK(I)</i>	The returns to one's rank in the educational attainment distribution for one's age-cohort are significant	WSH

In order to test $W(I)$ and $P(I)$ listed above, it is necessary to compare the RORs to education across different sectors of employment. However, selection into a specific employment sector is unlikely to be an arbitrary occurrence, but rather to be based on endogenous choice. The self-employed, for instance, may choose to be self-employed because of the appeal of owning their own business or because they do not have the opportunity to procure employment elsewhere (Koch and Ntege, 2008:17). Failure to account for such sample selection biases estimates and leads to unreliable inference in regression analysis.

The most frequently used method to control for endogenous choice is to conduct a Heckman two-step sample selection correction.¹⁰ While this approach gains tractability in the simplicity of its application, it explicitly allows only for binomial selection outcomes (Heckman, 1979:156). However, in the analysis that follows, five distinct employment selection outcomes are considered using the broad definition of the labour force: non-participation in the labour force, unemployment, self-employment, private employment and public employment.¹¹ Therefore, an adjusted version of the Dubin and McFadden (1984) method, as suggested by Bourguignon, Fournier and Gurgand (2007), is implemented in the analysis below to account for these multinomial selection effects.

In this regression analysis, Blacks, Whites, and Coloureds in South Africa are considered separately to test the hypothesis that sorting may be more prevalent for certain race groups. However, the disaggregation by race and sector of employment may potentially give rise to sample size issues. Therefore, the data set that is utilised for the analysis comprises a pooled sample of Statistics South Africa's March and September Labour Force Surveys (LFS) from 2001 to 2007. The pooling of this data set relies on the assumption that the data generating processes with regard to the hypotheses tested remained constant between 2001 and 2007. To account for potential differential effects between surveys, dummy variables for the LFSs are included in the regressions where applicable. Furthermore, the analysis considers only the population of working age (15 to 65 year olds).

¹⁰ See Heckman (1979).

¹¹ These five categories represent the mutually exclusive and collectively exhaustive choices for the population of working age. Therefore, the probabilities of being selected into each of these categories should collectively sum to one for every individual considered.

5.2 The Empirical Model

Similar to the Heckman (1979) procedure, the Dubin and Mcfadden (1984) method involves a two-step approach. In the first stage a multinomial logit model is estimated for selection into different sectors of employment. In the second stage, ordinary least squares is used to estimate a regression for the log of wages for employed individuals which includes the selection correction terms from the first-stage regression to account for the potential correlation of error terms between the respective sector of employment choices (Mansur, Mendelsohn and Morrison, 2008:177).

To explain this method, denote the private sector, public sector, self-employed sector, the unemployed, and the non-participating by p, g, s, u and n respectively. Furthermore, denote the log of wages by w and the discrete variable indicating the sector of choice by y . The aim is, for instance, to estimate the expected private sector wage $E(w_p|x, z)$, where x represents the exogenous variables influencing the log of wages and $x \subset z$ for parametric identification (Koch and Ntege, 2008:18). However, the log of wages in the private sector can only be observed for individuals that are employed in the private sector so that one can specify:

$$w_p = x\beta_p + \epsilon_p \quad \text{and}$$

$$y_j^* = zy_j + \eta_j, \quad \text{with } j = n, u, s, p, g$$

Where $E(\epsilon_p|x, z) = 0$ and $V(\epsilon_p|x, z) = \sigma^2$ (Bourguignon, Fournier and Gurgand, 2007:3). From this specification, it is clear that w_p is only observed if $y_p^* > \max_{j \neq p} y_j^*$ (Mansur, Mendelsohn and Morrison, 2008:178). This means that private sector wages can only be observed if an individual selects to be employed privately, which would be the case only if the “utilities” derived from private sector employment are greater than those derived from choosing any other sector in the set j (Bourguignon, Fournier and Gurgand, 2007:3). When the residuals, $\eta_i \dots \eta_k$, are independently and identically Gumbel distributed, the multinomial logit model is sufficient for estimating the discrete choice outcome. However, to consistently estimate the parameter β_p , it is necessary to account for the possibility that $E(\epsilon_g|x, z) \neq 0$. Therefore:

$$\begin{aligned} E(w_p|x, z) &= E\left(x\beta_p + \epsilon_p \mid y_p^* > \max_{j \neq p} y_j^*\right) \\ &= x\beta_p + E(\epsilon_p \mid y_p^* > y_n^*, y_p^* > y_u^*, y_p^* > y_s^*, y_p^* > y_g^*) \\ &= x\beta_p + \mu(P) \end{aligned}$$

Where $\mu(P)$ is the selection term for selection into the private sector of employment (Koch and Ntege, 2008:19). To correctly estimate the parameter of interest, β_p , it is necessary to account for any correlation between the explanatory variables and the disturbance term in the first-stage regression that is introduced when the disturbance term ϵ_p is not independent of all the η_j 's (Bourguignon, Fournier and Gurgand, 2007:4). Failure to do so would result in omitted variable bias such that the estimated coefficients would be biased. OLS estimation is therefore insufficient for this analysis and it is necessary to run a multinomial logit selection correction model, such as can be performed using the "SELMLOG" module in the statistical package STATA/SE 10.0 (Bourguignon, Fournier and Gurgand, 2002).

5.3 Empirical Specification and Results

Following the methodology discussed above, multinomial logit (MNL) selection correction regressions, disaggregated by sector of employment and testable hypothesis, are estimated for each of the race groups under consideration. All of the estimated regressions include age-cohort-, gender-, and LFS-period dummies. With the exception of the second-stage Mincerian earnings regression for the $P(II)$ and $SK(I)$ hypotheses, all regressions also include educational attainment dummies.¹² Unless stated otherwise, the reference category is 15 to 20 year-old males with no formal education in the March 2001 LFS.

Parametric identification of the second stage Mincerian earnings function requires that the first-stage MNL includes additional identification variables (Bourguignon, Fournier and Gurgand, 2007:3). To this end, a number of household level variables (number of children in the household, number of males and females aged 0 to 5, 6 to 12, and 13 to 18 in the household, and the number of elderly persons in the household) are included as exclusion restrictions in the MNLs. All the MNLs have identical specifications, but the comparison sector may be different for each race group.¹³ Furthermore, to test $P(II)$, it is necessary to include a variable for tenure in the second-stage regressions. However, the Tenure variable in the LFSs has missing values unless an individual is employed in the public or private

¹² These education dummies represent the following levels of educational attainment: primary education (Primary), incomplete secondary education (NCS), completed secondary education (Matric), tertiary diploma or certificate (Diploma), bachelors degree (Bachelors), and honours degree, bachelors degree with a diploma, masters degree or higher (Postgrad). The second—stage regression for the P(II) hypothesis includes a variable for individuals' years of completed education.

¹³ The sector with the most observations is automatically chosen by Stata/SE 10.0 as the comparison sector. Unlike other Stata/SE 10.0 modules, the add-on module 'SELMLOG' used here does not allow one to manually choose the comparison sector.

sectors.¹⁴ Consequently, $P(II)$ can only be tested for the private and public sectors and not for the self-employed. The sample on which the estimation for this hypothesis is based therefore excludes all individuals who do not meet these criteria.

The relevant first-stage MNLs are presented for each race group in Table 2 to Table 4 below. The coefficients in these regressions represent the discrete change in the log of odds for selection into a specific sector relative to the comparison sector and are therefore difficult to interpret. Consistent with theoretical prediction, it appears that education generally increases the probability of employment and lowers the probability of not participating in the labour market. These effects also appear to be stronger at higher levels of educational attainment.

At first glance, education seems to have an insignificant effect on the probability of self-employment for Whites, while having strong positive effects on the probability of self-employment for Blacks and Coloureds. However, since the comparison sector for the White race group is “private employment”, the insignificant coefficients for the self-employed regression simply denote that, for Whites, the probability of becoming self-employed is not statistically different from the probability of being privately employed.¹⁵

The second-stage MNL corrected Mincerian earnings regressions are presented in Table 5 to Table 7 below. To simplify the testing of $W(I)$ and $P(I)$, the marginal effects of educational attainment on the log of wages relative to the “no schooling” reference category for the public, private, and self-employed sectors are graphed for each race group from the $W(I)$, $P(I)$ regressions in Figure 1 to Figure 3 below. The corresponding upper and lower limits of the 95% coefficient confidence intervals are shown as dashed lines (matching the colours of the relevant sectors) on the graphs. Graphically, the returns to education (the percentage wage premium relative to no schooling) in one sector are statistically different from those in another sector if its coefficient nodes fall outside the confidence interval for the comparison sector.

In the figure for Blacks, the returns to education are consistently lower for the self-employed than for the privately and publicly employed. While the returns to education for Whites are also lower for the self-employed than the privately employed, the returns to education for the publicly employed lie within the 95% confidence interval for the self-employed sector (and

¹⁴ The Tenure variable in the LFS accounts for the number of years that an individual has been working for his/her current employer

¹⁵ The first-stage MNL results are not discussed in detail here since the relevant hypothesis all relate to the second-stage Mincerian earnings regressions.

Table 2 - Multinomial Logits for Black Individuals

Relevant Hypothesis	<i>W(I), P(I), P(II), SK(I)</i> ^a			
	Unemployed	Public	Self-Empl	Private
Primary	0.5754 (0.0135)***	0.4707 (0.0133)***	0.7241 (0.0321)***	0.4431 (0.0162)***
NCS	0.4619 (0.0137)***	0.3153 (0.0137)***	1.2737 (0.0314)***	0.3447 (0.0171)***
Matric	1.4599 (0.0164)***	1.2861 (0.0171)***	3.3133 (0.0331)***	1.0899 (0.0219)***
Diploma	1.2651 (0.0333)***	1.377 (0.0343)***	5.3921 (0.0422)***	1.1922 (0.0434)***
Bachelors	0.7592 (0.07)***	1.1574 (0.0688)***	5.3398 (0.0699)***	1.1448 (0.0848)***
Postgraduate	0.3866 (0.0961)***	1.1117 (0.0876)***	5.4447 (0.085)***	1.2135 (0.1048)***
<i>Number of Children in HH</i>	-0.0056 (0.0077)	-0.0355 (0.0103)***	0.0286 (0.0181)	0.0201 (0.0119)***
<i>Males aged 0 to 5</i>	0.1097 (0.0096)***	-0.0851 (0.0127)***	-0.0519 (0.0224)**	0.0487 (0.0147)***
<i>Females aged 0 to 5</i>	0.0991 (0.0095)***	-0.0981 (0.0127)***	-0.0491 (0.0223)**	0.021 (0.0147)
<i>Males aged 6 to 12</i>	0.0105 (0.0092)	-0.2028 (0.0123)***	-0.0702 (0.0213)***	0.0131 (0.0141)
<i>Females aged 6 to 12</i>	0.0161 (0.0092)***	-0.181 (0.0123)***	-0.0566 (0.0212)***	0.0148 (0.0141)
<i>Males aged 13 to 18</i>	-0.1284 (0.0063)***	-0.3433 (0.0084)***	-0.2206 (0.0149)***	-0.1002 (0.0097)***
<i>Females aged 13 to 18</i>	-0.0802 (0.0064)***	-0.2556 (0.0084)***	-0.1521 (0.015)***	-0.0886 (0.0099)***
<i>Number of Elderly in HH</i>	-0.0573 (0.0058)***	-0.8621 (0.0083)***	-0.7297 (0.0146)***	-0.4697 (0.0096)***
N	713098	713098	713098	713098
Pseudo R2	0.2262	0.2262	0.2262	0.2262
Log likelihood	-779274.41	-779274.41	-779274.41	-779274.41

Source: "SELMLOG" applied in STATA/SE 10.0.

Non-robust Standard Errors in Parenthesis. *** 1% Significance. ** 5% Significance. * 10% Significance.

a: Not Part of Labour Force is the comparison sector in multinomial logit for which all mnl parameters are normalized to zero

Table 3 - Multinomial Logits for Coloured Individuals

Relevant Hypothesis	<i>W(I), P(I), P(II), SK(I)</i> ^a			
	Not Partic	Unemployed	Public	Self-Empl
Primary	-0.1159 (0.0311)***	0.3695 (0.0418)***	0.4133 (0.0801)***	0.8518 (0.1082)***
NCS	-0.1126 (0.0314)***	0.2943 (0.0418)***	1.002 (0.0779)***	1.5689 (0.1058)***
Matric	-1.0303 (0.0388)***	-0.0284 (0.0447)	1.9605 (0.0791)***	1.6964 (0.1112)***
Diploma	-0.733 (0.0718)***	-0.4774 (0.0797)***	3.5834 (0.0851)***	2.0777 (0.1358)***
Bachelors	-0.6677 (0.1404)***	-1.0541 (0.1927)***	3.3645 (0.1133)***	2.357 (0.1842)***
Postgraduate	-0.7599 (0.191)***	-1.3137 (0.3045)***	3.4691 (0.1323)***	2.3134 (0.2251)***
<i>Number of Children in HH</i>	0.0543 (0.02)***	-0.0423 (0.021)**	-0.0056 (0.0365)	0.0072 (0.0503)
<i>Males aged 0 to 5</i>	-0.0728 (0.0253)***	0.1764 (0.0256)***	0.0431 (0.0445)	0.0091 (0.062)
<i>Females aged 0 to 5</i>	-0.0481 (0.0253)***	0.1558 (0.0256)***	0.0159 (0.0449)	-0.0524 (0.0628)
<i>Males aged 6 to 12</i>	0.0004 (0.0243)	0.1428 (0.0251)***	0.1469 (0.0429)***	0.0328 (0.0598)
<i>Females aged 6 to 12</i>	0.0117 (0.0242)	0.1235 (0.025)***	0.087 (0.043)**	0.0739 (0.0592)
<i>Males aged 13 to 18</i>	0.2081 (0.0166)***	0.101 (0.0174)***	0.0931 (0.0304)***	0.0176 (0.0418)
<i>Females aged 13 to 18</i>	0.0901 (0.0167)***	0.0762 (0.0173)***	0.0079 (0.0302)	-0.0059 (0.0414)
<i>Number of Elderly in HH</i>	0.4144 (0.0147)***	0.3562 (0.015)***	-0.0284 (0.0272)	0.0568 (0.035)
N	118758	118758	118758	118758
Pseudo R2	0.1615	0.1615	0.1615	0.1615
Log likelihood	-133301.76	-133301.76	-133301.76	-133301.76

Source: “SELMLOG” applied in STATA/SE 10.0.

Non-robust Standard Errors in Parenthesis. *** 1% Significance. ** 5% Significance. * 10% Significance.

a: Private sector is the comparison sector in multinomial logit for which all MNL parameters are normalized to zero

Table 4 - Multinomial Logits for White Individuals

Relevant Hypothesis	<i>W(I), P(I), P(II), SK(I)</i> ^a			
	Not Partic	Unemployed	Public	Self-Empl
Primary	-1.8245 (0.2861)***	-0.2151 (0.4546)	18.9639 .	-0.4555 (0.4578)
NCS	-2.3273 (0.2696)***	-0.5461 (0.4329)	19.35 (0.2576)***	-0.5367 (0.4331)
Matric	-2.9748 (0.2693)***	-1.0521 (0.4324)**	19.4931 (0.2559)***	-0.3402 (0.4325)
Diploma	-3.3331 (0.2708)***	-1.5331 (0.4353)***	20.1907 (0.2566)***	-0.1676 (0.4331)
Bachelors	-3.1213 (0.2721)***	-1.8625 (0.4409)***	20.4049 (0.2578)***	0.0349 (0.4337)
Postgraduate	-3.3651 (0.2743)***	-1.841 (0.4454)***	20.529 (0.2586)***	0.1084 (0.4342)
<i>Number of Children in HH</i>	0.1099 (0.0371)***	0.0205 (0.0564)	0.0771 (0.0524)	0.0567 (0.0446)
<i>Males aged 0 to 5</i>	-0.0565 (0.0497)	0.1037 (0.0711)	-0.1493 (0.0654)**	0.0969 (0.0545)***
<i>Females aged 0 to 5</i>	0.0046 (0.0515)	0.009 (0.0754)	-0.0064 (0.0658)	0.0466 (0.0566)
<i>Males aged 6 to 12</i>	-0.0649 (0.0472)	0.0801 (0.0699)	-0.0521 (0.0627)	-0.004 (0.0533)
<i>Females aged 6 to 12</i>	0.0528 (0.0479)	0.0658 (0.0717)	-0.0106 (0.0637)	0.0794 (0.054)
<i>Males aged 13 to 18</i>	0.1928 (0.0283)***	0.0775 (0.0441)***	0.0263 (0.041)	-0.0505 (0.0348)
<i>Females aged 13 to 18</i>	-0.0413 (0.0295)	0.0411 (0.0442)	-0.0411 (0.0417)	-0.0939 (0.0352)***
<i>Number of Elderly in HH</i>	0.4057 (0.0206)***	0.3933 (0.0333)***	0.0476 (0.0316)	0.0567 (0.0245)**
N	72636	72636	72636	72636
Pseudo R2	0.1586	0.1586	0.1586	0.1586
Log likelihood	-85905.106	-85905.106	-85905.106	-85905.106

Source: "SELMLOG" applied in STATA/SE 10.0.

Non-robust Standard Errors in Parenthesis. *** 1% Significance. ** 5% Significance. * 10% Significance.

a: Private sector is the comparison sector in multinomial logit for which all MNL parameters are normalized to zero

Table 5 – MNL Corrected Mincerian earnings functions (log of wages) for Black Individuals

Sector:	Self-Employed		Privately Employed		Publicly Employed	
	<i>W(I), P(I)</i>	<i>SK(I)</i>	<i>W(I), P(I)</i>	<i>P(II)</i>	<i>W(I), P(I)</i>	<i>SK(I)</i>
Primary	0.0712 (0.0259)***		0.2414 (0.0094)***		0.0735 (0.0258)***	
NCS	0.2557 (0.028)***		0.6029 (0.0098)***		0.4546 (0.0423)***	
Matric	0.4209 (0.0526)***		0.8953 (0.0179)***		0.7092 (0.087)***	
Diploma	0.0876 (0.1134)		1.4115 (0.034)***		0.9991 (0.1625)***	
Bachelors	0.8447 (0.1399)***		2.0595 (0.0422)***		1.2472 (0.1643)***	
Postgraduate	0.9543 (0.1551)***		2.0864 (0.0497)***		1.3466 (0.166)***	
Educ years				0.061 (0.0009)***	0.0919 (0.0025)***	
Tenure				0.0198 (0.0005)***	0.0456 (0.0013)***	
Educ years*Tenure				0.0011 (0.0001)***	-0.0023 (0.0001)***	
Rank		0.4634 (0.2899)				2.3852 (0.3481)***
Rank squared		-0.3911 (0.6416)				-3.8886 (0.7056)***
Rank cubed		0.5008 (0.4185)				2.8385 (0.4339)***
N	38141	38141	150856	150169	35655	35655
Adjusted R-sq	0.1286	0.1268	0.2689	0.3091	0.3580	0.3594

Source: “SELMLOG” applied in STATA/SE 10.0. Non-robust Standard Errors in Parenthesis. *** 1% Significance. ** 5% Significance. * 10% Significance.

Regressions also include: gender dummies, age-cohort dummies, and correction dummies from the first stage MNL. Complete results are available from the author, upon request.

Table 6 – MNL Corrected Mincerian earnings functions (log of wages) for Coloured Individuals

Sector:	Self-Employed			Privately Employed			Publicly Employed		
	<i>W(I), P(I)</i>	<i>SK(I)</i>		<i>W(I), P(I)</i>	<i>P(II)</i>	<i>SK(I)</i>	<i>W(I), P(I)</i>	<i>P(II)</i>	<i>SK(I)</i>
Relevant Hypothesis:									
Primary	0.4624 (0.2923)			0.3088 (0.0186)***			0.2013 (0.063)***		
NCS	0.7943 (0.518)			0.6578 (0.0254)***			0.6116 (0.0925)***		
Matric	1.0098 (0.5744)***			1.0216 (0.0331)***			1.0706 (0.1461)***		
Diploma	1.1293 (0.7719)			1.2648 (0.0706)***			1.4706 (0.2593)***		
Bachelors	1.8489 (0.878)**			1.6794 (0.0804)***			1.6701 (0.2513)***		
Postgraduate	1.5181 (0.8772)***			1.7368 (0.096)***			1.6963 (0.261)***		
Educ years					0.0862 (0.0021)***			0.1032 (0.0066)***	
Tenure					0.0058 (0.001)***			0.0403 (0.0028)***	
Educ years*Tenure					0.0016 (0.0001)***			-0.0015 (0.0003)***	
Rank		-1.7039 (1.2742)				1.951 (0.1318)***			0.9388 (0.47)**
Rank squared		3.6886 (2.7922)				-1.9252 (0.3326)***			-0.6002 (1.0434)
Rank cubed		-1.8793 (1.871)				0.932 (0.236)***			0.6508 (0.695)
N	2517	2517		42788	42522	42788	7165	7130	7165
Adjusted R-sq	0.2231	0.2244		0.3109	0.3467	0.3240	0.3960	0.4551	0.4046

Source: "SELMLOG" applied in STATA/SE 10.0. Non-robust Standard Errors in Parenthesis. *** 1% Significance. ** 5% Significance. * 10% Significance.

Regressions also include: gender dummies, age-cohort dummies, LFS survey dummies, and correction dummies from the first stage MNL. Complete results are available from the author, upon request.

Table 7 – MNL Corrected Mincerian earnings functions (log of wages) for White Individuals

Sector: Relevant Hypothesis:	Self-Employed		Privately Employed		Publicly Employed		
	<i>W(I), P(I)</i>	<i>SK(I)</i>	<i>W(I), P(I)</i>	<i>P(II)</i>	<i>W(D), P(I)</i>	<i>P(II)</i>	<i>SK(I)</i>
Primary	-0.1725 (0.5515)		0.902 (0.2038)***		<i>a</i>		
NCS	0.1068 (0.5342)		1.3083 (0.197)***		<i>a</i>		
Matric	0.1601 (0.5358)		1.3789 (0.1994)***		0.1359 (0.1663)		
Diploma	0.2376 (0.5427)		1.5384 (0.2041)***		0.2327 (0.177)		
Bachelors	0.5076 (0.548)		1.7155 (0.2069)***		0.5097 (0.2294)**		
Postgrad	0.5838 (0.5531)		1.8284 (0.2094)***		0.6963 (0.2508)***		
Educ years				0.1247 (0.0066)***		0.1245 (0.0133)***	
Tenure				0.0458 (0.0041)***		0.0377 (0.0068)***	
Educ years*Tenure				-0.0023 (0.0003)***		-0.0019 (0.0005)***	
Rank		3.149 (0.658)***			3.8984 (0.2718)***		3.3496 (0.5762)***
Rank squared		-6.9683 (1.354)***			-8.2437 (0.577)***		-7.047 (1.1634)***
Rank cubed		4.7129 (0.7845)***			5.1099 (0.3457)***		4.4316 (0.6418)***
N	7932	7932	23069	22892	5693	5666	5693
Adjusted R-sq	0.1068	0.1088	0.2551	0.2839	0.1843	0.2099	0.1888

Source: “SELMLOG” applied in STATA/SE 10.0. Non-robust Standard Errors in Parenthesis. *** 1% Significance. ** 5% Significance. * 10% Significance.

Regressions also include: gender dummies, age-cohort dummies, LFS survey dummies, and correction dummies from the first stage MNL. Complete results are available from the author, upon request.

a: Variables automatically dropped in estimation due to insufficient observations.

Figure 1 - Returns to Education for Black Individuals

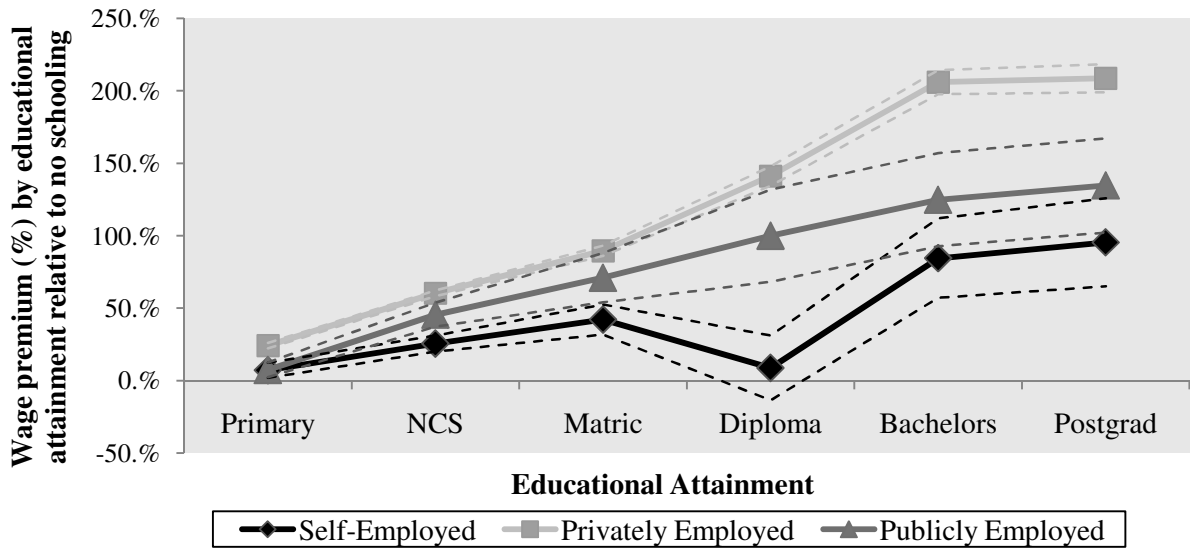


Figure 2 - Returns to Education for Coloured Individuals

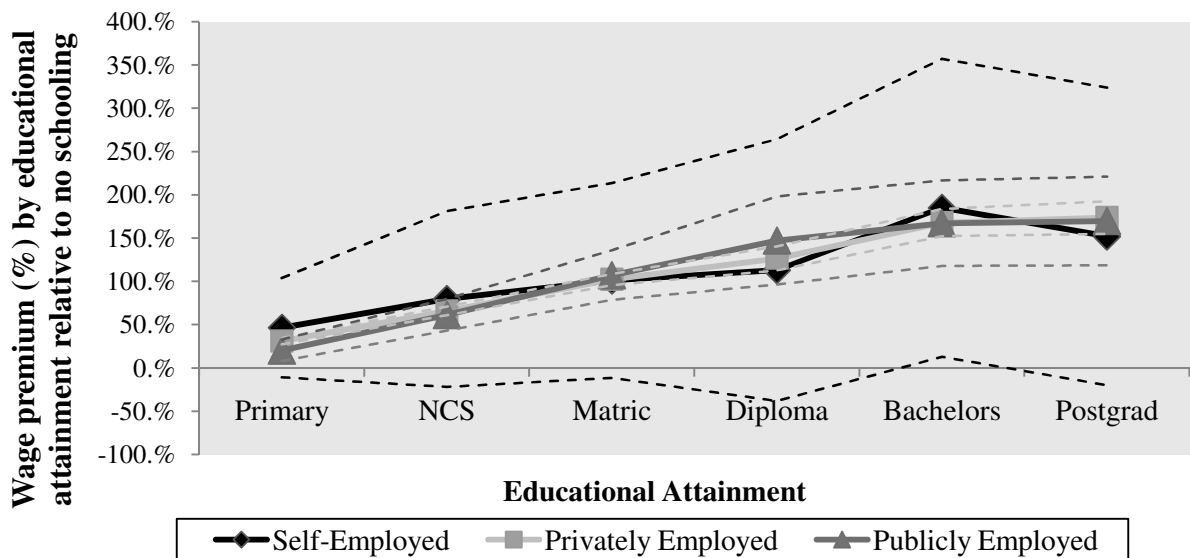
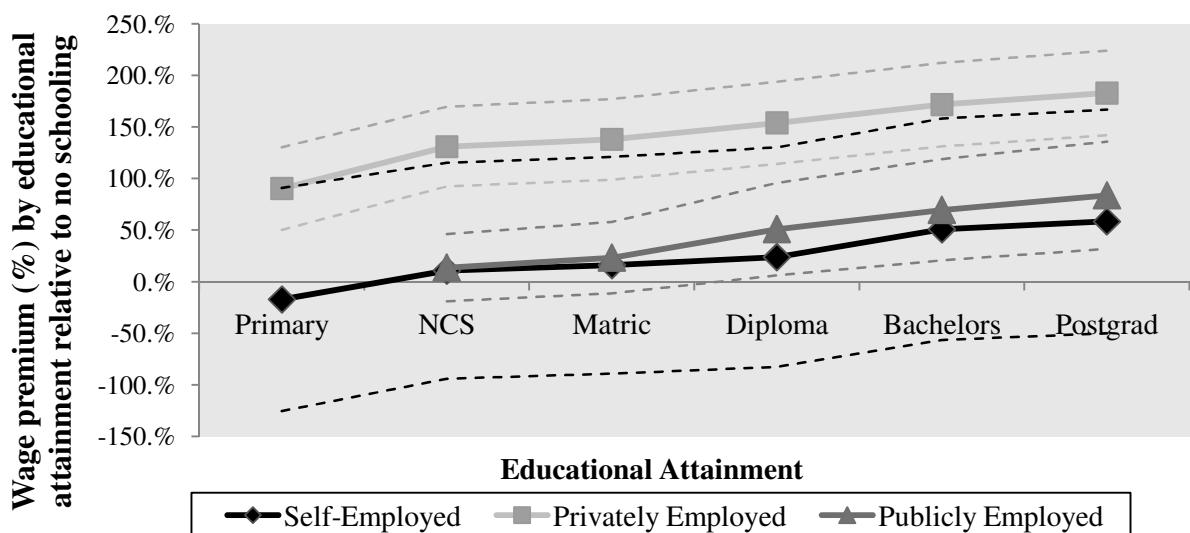


Figure 3 - Returns to Education for White Individuals



vice versa), denoting that the returns to education in these two sectors are not statistically different for this race group. For the Coloured race group, the returns to education in both the public and private sectors lie within the 95% confidence interval for the self-employed which shows that, generally, the returns to education are statistically the same for Coloureds in all three sectors. Overall, the figures thus suggest that $W(I)$ is confirmed for Blacks in the private and public sectors and for Whites in the private sector, but that it is refuted for Coloureds in both the private and public sectors.

From Tables 6 and 7 it must be noted that education seems to have a sporadic effect on the wages for self-employed Coloureds, and an overall insignificant effect on the wages for self-employed Whites. This result is consistent with the findings of Koch and Ntege (2008) who find mostly insignificant coefficients for educational attainment for the self-employed in South Africa and has two important implications: Firstly, it refutes Wolpin's (1977) assertion that the self-employed may invest in education purely for its productivity-augmenting function and secondly, it suggests a refutation of the HCT for self-employed Coloureds and Whites.

Figure 1 to Figure 3 indicates that the relationship between the returns to education in the public and private sectors varies by race-group and also somewhat by level of educational attainment. Since returns to education in the public sector in excess of those in the private sector would confirm $P(I)$, $P(I)$ is refuted for Blacks at all levels of educational attainment, but most strongly at the Bachelors and Postgraduate qualification levels. Similarly, $P(I)$ is strongly refuted for Whites at all levels of educational attainment. For Coloureds, $P(I)$ can only be confirmed at the Diploma qualification level and is otherwise refuted. Thus, Figures 1 to 3 seem to suggest that $P(I)$ is broadly refuted for all race-groups.

To test $P(II)$ for privately and publicly employed individuals, the approach of Altonji and Pierret (2001) is followed by including in the $P(II)$ regressions variables for the years of schooling, tenure, and the interaction between tenure and the number of years of completed formal education. A non-zero coefficient on the interaction-variable would indicate that the returns to education change as employers observe individuals' on-the-job productivities and would thus refute the SSH. The relevant estimated interaction-coefficients are negative for all races in all sectors excluding Blacks and Coloureds in the private sector for whom it is actually positive. Therefore, $P(II)$ and subsequently the SSH, is universally refuted for this sample.

Finally, in the $SK(I)$ regressions, the educational dummy variables are substituted by variables representing individuals' ranks in the educational attainment distributions for their age-cohorts.¹⁶ The overall effect of this rank is positive and significant for all groups except self-employed Black and Coloured individuals. This result, coupled with the significant coefficients to education for self-employed Black individuals in the $W(I)$, $P(I)$ regression, suggests that self-employed Blacks invest in education for its productivity-augmenting function in accordance with the HCT. Similarly, the finding that "rank" has a positive effect on wages for self-employed Whites corroborates the finding from the $W(I)$, $P(I)$ regressions that self-employed whites appear to invest in education for its innate productivity-reflecting function. It would therefore appear that some White, self-employed individuals do indeed need to signal their abilities, perhaps not to employers, but to clients or business-partners. The private sector rank-effect also appears to be strongest for White individuals, indicating that sorting in the private sector may be most prevalent for this race-group.

The findings from the hypothesis tests discussed above are summarised in Table 8 below. The overall findings indicate that education in South Africa performs both a productivity-augmenting and a productivity reflecting function. While no evidence is found to support the SSH, the extent of sorting in the labour market appears to vary somewhat between sectors of employment and race-groups. Specifically, sorting seems to be most prevalent in the public sector for Blacks, most prevalent in the Private sector for Whites, and least prevalent for Coloureds overall.

Table 8 – Summarised findings from empirical tests of the SH in the South African labour market

	Blacks	Coloureds	Whites
Hypothesis	Confirmed for	Confirmed for	Confirmed for
$W(I)$	All levels of education in the Private and Public sectors	None	All levels of education in the Private sector
$P(I)$	None	Diploma	None
$P(II)$	No Sector	No Sector	No Sector
$KS(I)$	Private and Public Sectors	Private and Public Sectors	All Sectors

¹⁶ The "Rank" variable is constructed by calculating the frequency-weighted cumulative density function of educational attainment for each race group by age-cohort. Each individual is then assigned a value between 0 and 1 that denotes his/her rank within this cumulative density function. Quadratic and cubed terms for this variable are also included to allow for potential non-linearity in its effects.

5 CONCLUSION

This paper has examined some of the important concepts and issues pertaining to the Human Capital Theory (HCT) and the Sorting Hypothesis as they relate to education in the South African labour market. The descriptive analysis considered the relevance of the HCT-SH debate and discussed some general and specific implications of the two theories, specifically relating to who should finance education under certain circumstances. These implications revealed the importance of determining the extent to which education augments productivity as opposed to reflecting innate productivity in South Africa.

In the empirical analysis, the methods of Wolpin (1977), Psacharopoulos (1979), and Kroch and Sjoblom (1994) were used in an attempt to determine the validity of the SH for different race-groups employed in different sectors in the labour market. No evidence was found to suggest that education performs an exclusively productivity-reflecting function and the SSH is subsequently refuted for the South African labour market. However, some evidence for the Weak Sorting Hypothesis was found for all race groups. While self-employed Blacks appear to invest in education for its productivity-enhancing function, their education is found to perform a considerable sorting function in the private and public sectors of the economy. It is found that education performs a significant sorting function for Whites employed in the private sector, particularly considering the fact that self-employed Whites also appear to invest in education for its innate productivity-reflecting function. Lastly, the hypothesis tests for Coloureds rendered some conflicting results, but appear to suggest that sorting matters much less for Coloureds than for Blacks or Whites.

The findings in this paper thus suggest that education in South Africa acts as both an augments of productivity and a reflector of innate productivity. From an education policy perspective, these findings suggest that the proportion of education costs that should be borne privately should be greatest for Whites and that public funding of education should be greatest for Coloureds. However, while an attempt was made to control for differences in the extent of sorting across race groups, the magnitudes of the HCT and SH components of the labour market returns to education still cannot be estimated using the approaches applied here. Consequently, in the knowledge that sorting considerations matter in the South African labour market, the next step should be to investigate which industries and occupations are likely to screen individuals on the basis of educational attainment and, similarly, which education credentials are commonly acquired as signals of innate ability and not for their

human capital-enriching function. While such an approach still may not reveal the magnitudes of the HCT and SH components of education, it could identify those fields of study that deserve priority in public financing.

Finally, a potential future direction for tests of the SH is to switch focus from the earnings-returns to education and instead investigate the rate of return to education with regard to the probability of procuring employment. After all, the theoretical predictions of the HCT and the SH are entirely consistent with the manifestation of sorting in terms of education's role in raising the probability of employment. In a country characterised by high levels of joblessness, such an approach may be indeed be more appropriate and could, potentially, provide new insights into education's role as a sorting mechanism. Consequently, the development of further methods for testing the extent of sorting are imperative for establishing which theory and, subsequently, which set of policy prescriptions with respect to education is more applicable in South Africa.

REFERENCES

- ALTONJI, J., and PIERRET, C. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116 (1), 313-350.
- ARABSHEIBANI, G. R., and REES, H. (1998). On the Weak vs Strong version of the Screening Hypothesis: A Re-Examination of the P-Test for the U.K. *Economics of Education Review* , 17 (2), 189-192.
- ARROW, K. J. (1973). Higher Education as a Filter. *Journal of Public Economics*, 2, 193-216.
- ASHENFELTER, O., and KRUEGER, A. (1992). Estimates of the Economic Return to Schooling from a New Sample of Twins. *The American Economic Review* , 85 (5), 1157-1173.
- BECKER, G. S. (1992a). Human Capital and the Economy. *Proceedings of the American Philosophical Society* , 136 (1), 85-92.
- BECKER, G. S. (1992b). The Economic Way of Looking at Life. *Nobel Lecture* 9 December 1992. Oslo, Norway.

- BERG, I. (1970). *Education and Jobs: The Great training Robbery*. New York: Praeger Publishers.
- BHORAT, H., and McCORD, A. (2003). Employment and labour market trends in South Africa, in HSRC (ed.) *Human Resources Development Review 2003*, HSRC Press, Cape Town.
- BLAUG, M. (1985). Where are we now in the economics of education? *Economics of Education Review*, 4 (1), 17-28.
- BOURGUIGNON, F., FOURNIER, M., and GURGAND, M. (2002). SELMLOG: Stata module to perform selection bias correction based on the multinomial logit model. [online] available from <http://www.crest.fr/pageperso/lmi/gurgand/selmlog.htm>. [10 November 2008]
- BOURGUIGNON, F., FOURNIER, M., and GURGAND, M. (2007). Selection bias corrections based on the multinomial logit model: Monte-Carlo comparisons. *Journal of Economic Surveys*, 21 (1), 174-205.
- BROWN, S., and SESSIONS, J. G. (1999). Education and employment status: a test of the strong screening hypothesis in Italy. *Economics of Education Review*, 18, 397-404.
- BROWN, S., and SESSIONS, J. G. (1998). Education, Employment Status and Earnings: A Comparative Test of the Strong Screening Hypothesis. *Scottish Journal of Political Economy*, 45 (5), 586-591.
- BROWN, S., and SESSIONS, J. G. (2005). Evidence on the relationship between firm-based screening and the returns to education. *Economics of Education Review* (25), 498-509.
- BURGER, R., and JAFTA, R. (2006). *Returns to Race: Labour Market Discrimination in Post-Apartheid South Africa*. Stellenbosch Economic Working Papers No. 04/06. Department of Economics, University of Stellenbosch.
- CASTAGNETTI, C., CHELLI, F., & ROSTI, L. (2005). Educational Performance as Signalling Device: Evidence from Italy. *Economics Bulletin*, 9 (4), 1-7.
- CHATTERJI, M., SEAMAN, P. T., and SINGELL, L. D. (2003). A test of the signalling hypothesis. *Oxford Economic Papers*, 55, 191-215.

- CLARK, A. (2000, March). *Signalling and Screening in a Transition Economy - Three Empirical Models Applied to Russia*. Discussion Paper No 2000/03, Centre for Economic Reform and Transformation, Department of Economics, Heriot-Watt University, Edinburgh.
- COHN, E., and GESKE, T. G. (1990). *Economics of Education*. Pergamon Press: Oxford.
- DUBIN, J. A., and MCFADDEN, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* , 52 (2), 345-362.
- GROOT, W., and OOSTERBEEK, H. (1994). Earnings Effects of Different Components of Schooling; Human Capital versus Screening. *The Review of Economics and Statistics* , 317-321.
- HASPEL, A. E. (1978). The Questionable Role of Higher Education as an Occupational Screening Device. *Higher Education* , 7 (3), 279-294.
- HECKMAN, J. (1979). Sample selection bias as a specification error. *Econometrica* , 47 (1), 153-161.
- HEYWOOD, J. S., and WEI, X. (2004). Education and Signalling: Evidence from a Highly Competitive Labor Market. *Education Economics* , 12 (1), 1-16.
- JAAG, C. (2006). *Institutions in Education and the Importance of Skill Formation*. Unpublished Phd Dissertation, University of St. Gallen, Zurich .
- KOCH, S. F., and NTEGE, S. S. (2006a). Education Screening in South Africa's Labour Market? *DPRU Conference on Accelerated and Shared Growth in South Africa: Determinants, Constraints and Opportunities*. held 18 - 20 October 2006, Johannesburg, South Africa.
- KOCH, S. F., and NTEGE, S. S. (2006b). Potential Screening in South Africa's Labour Market. *SARB Conference 2006* , 143 - 150.
- KOCH, S. F., and NTEGE, S. S. (2008). *Returns to Schooling: Skill Accumulation or Information Revelation?* University of Pretoria, Department of Economics Working Paper Series Working Paper No. 2008-12.
- KROCH, E. A., and SJOBLUM, K. (1994). Schooling as Human Capital or a Signal: Some Evidence. *The Journal of Human Resources* , 29 (1), 156-180.

- LAYARD, R., and PSACHAROPOULOS, G. (1974). The Screening Hypothesis and the Returns to Education. *Journal of Political Economy* , 82 (5), 985-998.
- LIU, L. W., and WONG, Y. C. (1982). Education and Screening: A Test. *Economic Inquiry* , 20 (January), 72-83.
- MAGOULA, T., and PSACHAROPOULOS, G. (1999). Schooling and monetary rewards in Greece: an over-education false alarm? *Applied Economics* (31), 1589-1597.
- MANSUR, E. T., MENDELSON, R., and MORRISON, W. (2008). Climate change adaptation: A study of fuel choice and consumption in the US energy sector. *Journal of Environmental Economics and Management* (55), 175-193.
- PSACHAROPOULOS, G. (1979). On the weak versus the strong version of the screening hypothesis. *Economics Letters* (4), 181-1855.
- PSACHAROPOULOS, G. (1995). *The Profitability of Investment in Education: Concepts and Methods*. Human Capital Development Operations Policy Working Paper.
- QUIGGIN, J. (1999). Human capital theory and education policy in Australia. *Australian Economic Review* , 32 (2), 130-44.
- RILEY, J. G. (1979). Testing the Educational Screening Hypothesis. *Journal of Political Economy* , 87 (5), 227-251.
- SAKAMOTO, A., and KIM, C. (2005). *Estimating the Human Capital and Screening Effects of Schooling on Productivity in U.S. Manufacturing Industries, 1979-1996*. [online] available from <http://paa2006.princeton.edu/download.aspx?submissionId=60538>. [20 July 2008]
- SPENCE, M. (1973). Job Market Signalling. *The Quarterly Journal of Economics* , 87 (3), 355-374.
- STIGLITZ, J. E. (1975). *The Theory of "Screening", Education, and the Distribution of Income*. *The American Economic Review* , 65 (3), 283-300.
- STIGLITZ, J. E., and WEISS, A. (1990). *Sorting out the Differences between Signalling and Screening Models*. NBER Technical Working Paper No. 93.
- TAUBMAN, P. J. and WALES, T. J. (1973). Higher Education, Mental Ability, and Screening. *Journal of Political Economy* , 81 (1), 28-55.

- VAN DER BERG, S. (2005). Apartheid's enduring legacy: Inequalities in education. Paper delivered to the Oxford University/University of Stellenbosch conference on *The South African Economic Policy under Democracy: A 10 year review*, Stellenbosch, October 2005.
- WEISS, A. (1995). Human Capital vs Signalling Explanation of Wages. *The Journal of Economic Perspectives*, 9 (4), 133-154.
- WOLPIN, K.I. (1977). Education and Screening. *American Economic Review*, 67 (5), 949-958