IMPACT OF HEALTH ON LABOUR FORCE PARTICIPATION IN SOUTH AFRICA[1]

Chijioke O. Nwosu[a,b]

Ingrid Woolard[b]

ABSTRACT

This paper quantifies the impact of health on labour force participation, using South Africa as a case study. This is important given the essential role the labour market plays in economic growth and the potential for poor health to adversely affect labour market outcomes. South Africa has experienced significant disease burden especially due to communicable diseases like HIV/AIDS and tuberculosis. Moreover, conditions like obesity remain a public health concern. Furthermore, the country has witnessed declining labour force participation in recent years. These health and labour market outcomes, coupled with relatively scant literature on the impact of health on the labour market in South Africa, motivate this study. Data is sourced from the first and third waves of the National Income Dynamics Study, a nationally representative panel dataset of South African households and a rich source of health and socio-economic data. Endogenous treatment of self-assessed health in a contemporaneous setting suggests positive and significant impact of health on labour force participation. The hypothesis of exogeneity of self-assessed health in a labour force participation equation is however not rejected. Finally, positive and significant association between health and LFP persists even four years after health assessment.

Keywords: Labour force participation; Health; Instrumental variables; Average treatment effect; Treatment effect on the treated; Local average treatment effect.

JEL codes: I15; J21

[a] Department of Economics, Andrew Young School of Policy Studies, Georgia State University, Atlanta, USA. Corresponding author: nwosuchijioke@gmail.com

[b] School of Economics, University of Cape Town, South Africa

[1] The authors are grateful to Patrizio Piraino and an anonymous referee for invaluable comments on earlier drafts of this paper. All remaining errors are attributable to the authors. C.O.N. gratefully acknowledges financial support from the Carnegie Corporation and the African Economic Research Consortium.

INTRODUCTION

The labour market is an important institution which mediates the relationship between health and the economy. Jack and Lewis (2009) maintain that the most obvious reason why healthier people are more likely to be richer than their sicker counterparts is that they have a greater capacity to work harder, longer and more consistently. Poor health therefore generally affects productivity and output adversely, thereby discouraging the sick from participating in productive activities due to decreased opportunity cost of leisure. Apart from reduced productivity, ill health may increase the utility derived from time spent away from market-related activities given that seeking health care is time-consuming. Conversely, low earnings due to poor health may induce an income effect, thereby resulting in increased labour supply (an effect which may be reinforced by often high cost implications of ill health); this latter effect suggests a negative relationship between better health and labour market participation, though it is more common to hypothesize a positive relationship (Cai & Kalb, 2006). It is therefore important to quantify the impact of health on labour market participation especially in developing countries where physical fitness is a very important asset given the mainly manual nature of most jobs.

However, there are complications in establishing the impact of health on labour force participation (LFP) given an apparent simultaneity between both health and participation. This is because though health increases people's ability/willingness to participate, higher incomes and self-esteem from gainful participation will likely improve health status. Also, participation may result in sickness for the employed due to stress and poor working conditions. However, the boredom associated with non-participation may result in poor health outcomes. These relationships possibly imbue health with endogeneity which complicates the analysis of the impact of health on participation (Cai & Kalb, 2006).

Furthermore, virtually every health measure suffers from measurement error in labour market equations as it is very difficult to obtain a health measure that perfectly reflects work capacity (Currie & Madrian, 1999). This leads to a bias similar to the ability bias in standard human capital models (Griliches, 1977), thereby resulting in attenuation bias (Cameron & Trivedi, 2005). Though some studies argue that more comprehensive health measures (e.g. self-assessed health (SAH)) increase the explanatory power of health in labour supply models over single indicator health measures (Manning, Newhouse, & Ware Jr, 1982), others prescribe the use of relatively objective health measures (e.g. life expectancy and limitations with activities of daily living) over more subjective measures like SAH (Kreider, 1999). This is because the measurement error associated with self-reported health might not be random, where non-labour force participants may be more likely than participants to cite illness as the reason for their non-participation given the social stigma associated with non-participation and the fact that the receipt of certain public transfers is dependent on health status (Boskin, 1977; Currie & Madrian, 1999; Parsons, 1980). This potentially leads to a bias referred to as rationalization endogeneity, likely to result in an over-estimation of the impact of health and the under-estimation of the effect of financial variables on labour supply (Bound, 1991; Bound et al., 1995; Cai, 2010; Cai & Kalb, 2006). But to the extent that these objective health measures are imperfect measures of true health status, they are likely to under-estimate the impact of health on labour supply.

Given the foregoing, Bound (1991) suggests that subjective health measures might be associated with less bias than objective ones since they would likely be affected by two opposite sources of bias (rationalization endogeneity and error-in-variables) which might cancel out, while objective health measures are only likely to be biased downwards (due to errors-in-variables bias). Furthermore, SAH, has been shown to emanate from respondents' rational thought processes while encompassing various dimensions of their health (including cultural and biological), as well as bodily sensations which may not be easily detected via clinical tests (Jylhä, 2009). Little wonder, it predicts mortality well (Ardington & Gasealahwe, 2014). These facts have bolstered the support for SAH as a useful health indicator.

Though many studies have examined the trends and determinants of labour supply in South Africa, the possible impact of health has been largely ignored (Banerjee, Galiani, Levinsohn, McLaren, & Woolard, 2008; Bhorat, 2007; Ntuli & Wittenberg, 2013; Wittenberg, 1999). Given that available data show that prime-age LFP has been declining in recent years in South Africa and the important role that gainful labour market participation plays in economic growth, it becomes imperative to investigate factors responsible for a decline in participation. While the strict LFP rate increased from 51.4% to 59.4% between 1995 and 2001, it declined to 57.2% in 2005 (Banerjee et al., 2008), with a further decline to 54.3% in the final quarter of 2011 (Statistics South Africa, 2012a).

Furthermore, the country is beset with substantial disease burden as a result of the high incidence of the HIV/AIDS pandemic and rising morbidity from non-communicable diseases (NCDs). Indeed, South Africa has been described as suffering from a quadruple burden of disease due to mortality from communicable diseases, maternal and perinatal conditions and nutritional deficiencies; NCDs like diabetes and obesity; injuries; and HIV/AIDS[2] (Bradshaw et al., 2000). As Figure 1 shows, South Africa had a high disease burden in 2004 even relative to other developing countries. And compared to 2000 disability-adjusted life year (DALY) figures (approximately 16 million), 2004 figures (approximately 22 million[3]) represent a substantial increase in DALYs even in the face of (at least slightly) declining HIV/AIDS prevalence (AVERT, undated), suggesting increased morbidity from non-HIV/AIDS-related causes.

[Figure 1]

Given the foregoing, this study estimates the impact of health on LFP in South Africa using bivariate probit and instrumental variables (IV) models. We find positive and statistically significant impact of better health on LFP for both males and females, where the effect is more pronounced among males. This study is timely given the paucity of empirical evidence on the impact of health on LFP in South Africa.

Several studies have been conducted on the health-labour supply relationship especially in developed countries. Stern (1989) examined the simultaneous impact of health (health limitations and SAH) on LFP and vice versa in the US using the 1978 Survey of Disability and Work and the 1979 cohort of the Health Interview Survey. The health equations and labour participation equations were identified by excluding marital status and its interaction with sex from the former, and various health conditions (like blindness, weakness and walking problems) from the latter. Results showed that self-reported disability status significantly increased LFP while there was only evidence of weak endogeneity of disability on LFP. Also, there was no evidence of systematic over-reporting of disability by non-labour force participants. However, a drawback of the study is the difficulty in justifying the exclusion of marital status from the health equation as well as inadequate test of the exogeneity of SAH. However, while showing that better SAH increased the probability of LFP for all sex-age

[2] Though HIV/AIDS conventionally belongs to the communicable disease group, its unusually high prevalence in South Africa led to its separate classification.
[3] Calculated from the South African DALY in Figure 1 and a total population of 48 million.

groups, similar studies conducted in Australia by Cai and Kalb (2006) and Cai (2010) rejected the hypothesis of SAH exogeneity.

Other studies that have used health conditions and/or some measure of physical functioning as SAH instruments include Dwyer and Mitchell (1999) and Campolieti (2002). The case for using them as instruments for SAH stems from human capital theory where health is seen as a form of human capital. Thus, it is taken that adverse health conditions/limitations in physical functioning affect participation in the labour market only through their effects on the underlying health capital (here proxied by SAH).

Using the Health and Retirement Study data, Benitez-Silva et al. (2004) found that disability applicants did not over-state their disability status on average. Even though Au et al. (2005) found some evidence of non-labour force participants using illness to rationalize their labour market status in Canada, employment effects of self-reported health were similar to what obtained when more objective health measures were used. Also, Campolieti (2002) found a large negative effect of disability on the LFP of older Canadian males but suggested that using self-reported health measures will under-estimate the impact of disability on labour force decisions.

Though O'Donnell (1998) suggested that non-workers' disability responses were unreliable, he was of the view that such a conclusion should only be made after conducting formal tests. In his work on the effect of disability on employment among the disabled in the UK, he made the case for recognition that disability potentially leads to work incapacity. He contended that some people were too disabled to work at all. This is supported by Nagi (1969) who observed that 55% of US disability insurance applicants were not fit for any kind of work, as well as Bound (1989) who concluded that many disability insurance applicants would not be working even in the absence of the disability insurance programme because of the extent of their disability.

Gomez and Nicolas (2006) employed difference in difference and matching techniques in studying the simultaneous causal relationship between health (SAH) shocks and employment in the Spanish labour market using the European Community Household Panel. They found that individuals who suffered a health shock were around 5% less likely to remain in employment and 3.5% more likely to remain inactive than those who did not suffer any health shock.

Haan and Myck (2009) examined the dynamic (sequential) simultaneous relationship between health (SAH) shocks and non-employment risk of adult males aged 30-59 years using the 1996-2007 waves of the German Socio-economic Panel. Similar to Bartel and Taubman (1986) and Haveman et al. (1994), lagged health status (non-employment) was used to predict current non-employment risk (health status) in order to avoid endogeneity. Results showed that lagged SAH (non-employment risk) exerted a significant causal impact on current non-employment (SAH) whether or not unobserved heterogeneity was controlled for.

Some studies have tracked the relationship between health status and future labour market outcomes. Using the Panel Study of Income Dynamics (PSID), Smith (2009) found that having self-reported one's health to be excellent or very good up to age 16 was associated with a positive and statistically significant effect on future hours of work relative to worse health outcomes. Also in a study that explored own/sibling's labour market implications of health outcomes using the relatively income poor sub-sample of the PSID, Choi (2007) found that an unhealthy young adult was less likely to work as an adult, while females with unhealthy siblings in young adulthood were more likely to work as adults relative to other females, thus suggestive of long term effects of illness as well as intra-family income transfer requirements.

With regard to developing country evidence, Mete and Schultz (2002) found a negative association between poor health and LFP among the elderly in Taiwan. Bridges and Lawson (2008) also found that poor health was associated with a decline in the probability of being in the formal labour market in Uganda, where this relationship was stronger among women. Moreover, investigating the relationship between health status and LFP in Sub-Saharan Africa using a dynamic panel data model with 46 countries, Novignon, Novignon and Arthur (2015) found a positive and significant relationship between population health and LFP in the general and female populations.

Relatively few studies have investigated the health-labour market relationship in South Africa. They include Arndt and Lewis (2001), Booysen et al. (2002), Young (2005) and Levinsohn et al. (2013). While Arndt and Lewis found no significant impact of HIV/AIDS on employment in South Africa, Young found a significant long run relationship. Booysen et al. also found some relationship between the disease and labour supply in the Free State Province. Apart from not utilizing nationally representative survey data, these studies largely ignored the causal dimension of labour supply determination in South Africa. These concerns were addressed by Levinsohn et al. who found, using propensity scores, that being HIV-positive was associated with a 6-7 percentage point increase in unemployment probability. A common feature of these studies is that none evaluated the impact of a composite/global health indicator like SAH on these labour market outcomes, as each focused on a single health measure, mainly HIV/AIDS. Furthermore, we use IV techniques to ascertain the impact of health on LFP as opposed to the non-causal analysis done by most of these studies.

MODELS AND DATA

The endogeneity of SAH and the ability to recover estimates of average treatment effect (ATE), average effect of treatment on the treated (TOT) and the local average treatment effect (LATE) from the bivariate probit model (Angrist & Pischke, 2009) necessitated the estimation of the bivariate probit model. But given that it depends on the normality of the joint distribution of the error terms for consistency, the IV linear probability model (IV-LPM) was also estimated. However, two shortcomings of the LPM are heteroscedasticity and probability predictions outside the unit interval. In this application, the former is not a major problem given that all estimates were corrected for heteroscedasticity. Also, most of the predicted probabilities in the model lie within the unit interval; indeed, there was no negative predicted probability while only 2.4% of predictions exceeded 1. Among those that exceeded 1, the average value was only 1.02 while the largest value was 1.07, apparently suggesting that these predictions are not likely to exert significant influence on the estimates. IV-LPM has however, been shown to be quite robust and simple in implementation and interpretation (Angrist & Pischke, 2009). Given the nature of the bivariate probit model, the covariate of interest, SAH (five categories) was dichotomized, where the excellent, very good and good categories were grouped together while the fair and poor categories were coalesced. Figure 2 shows that this dichotomization is not unreasonable especially for women. To ascertain whether this “arbitrary” classification influenced the findings, another IV-LPM was estimated with the original five-level SAH (see Table 6 below).

Given the difficulty in obtaining convincing instruments, a number of other models were also estimated to ascertain the existence of a relationship, however non-causal, between health and LFP in both contemporaneous and temporal settings (see below). These include estimating the relationship between lagged health, negative health shocks and health improvement and LFP. Moreover, given the apparent gender bias in labour market outcomes in South Africa (see Figure 2), the analysis was disaggregated by gender.

In the first stage, we assume that latent health status (h^*) is a linear function of a vector of exogenous variables, X (e.g. socio-economic status) and excluded instruments (z) necessary to identify the impact of health on LFP, and a random error term (ε_H). In the second stage, unobserved latent LFP (l^*) is specified as a function of a vector of exogenous variables (X), latent health status (h^*) and a random error term, ε_L. Though z is an instrument vector, it is represented as a scalar in equation [1] below for notational simplicity. Importantly, there are possibly some unobserved joint determinants of both latent LFP and health status (captured by ρ below) such as genetic heterogeneity.

The bivariate probit model is specified as follows:

$$h_i^* = X^/\beta^* + \gamma^* z_i + \varepsilon_{i,H} \quad [1]$$

$$l_i^* = X^/\alpha^* + \delta^* h_i^* + \varepsilon_{i,L} \quad [2]$$

where $(\varepsilon_L, \varepsilon_H)$ is distributed bivariate normal with the following assumptions: $E(\varepsilon_{iL}) = E(\varepsilon_{i,H}) = 0; var(\varepsilon_{i,L}) = var(\varepsilon_{i,H}) = 1; corr(\varepsilon_{i,L}, \varepsilon_{i,H}) = \rho$ [see section 4.6.3 of Angrist and Pischke (2009) and section 15.7.3 of Wooldridge (2002)]. If $\rho = 0$, it will be indicative that SAH is exogenous.

Given that there are only discrete measures of both LFP and health, the above model was implemented via a non-linear transformation of the linear index model in equations [1] – [2] as follows:

$$h_i = 1[X^/\beta^* + \gamma^* z_i + \varepsilon_{i,H} > 0] \quad [3]$$

$$l_i = 1[X^/\alpha^* + \delta^* h_i^* + \varepsilon_{i,L} > 0] \quad [4]$$

From equation [4], an individual is deemed to participate in the labour force if her underlying latent LFP index exceeds zero. An analogous interpretation holds for health status as shown in equation [3]. Model [1]-[4] is similar to that used in modelling the impact of fertility on labour supply in the US (Angrist & Pischke, 2009) and the relationship between health status and poverty in South Africa (Godlonton & Keswell, 2005), where similar identification issues arise.

Model [1] – [4] was estimated as follows:

$$\Pr(h = 1|X, Z) = \Phi(X^/\beta + \gamma z_i + \varepsilon_{i,H}) \quad [5]$$

$$\Pr(l = 1|X) = \Phi(X^/\alpha + \delta h_i + \varepsilon_{i,L}) \quad [6]$$

where for instance, the joint probability of a respondent being in SAH=1 and LFP=1 (using the above specifications) is:

$$\Pr(h = 1, l = 1) = \Phi_b(X^/\beta + \gamma z_i, X^/\alpha + \delta h_i, \rho)$$

where Φ_b is the bivariate normal cumulative distribution function (CDF) (Cameron & Trivedi, 2005). A formal test of exogeneity entails testing whether the correlation coefficient, ρ is statistically different from zero. If ρ is statistically significant, it is evidence of the endogeneity of SAH. Statistical insignificance of ρ would suggest that estimating separate LFP and SAH equations may not produce inconsistent estimates. The IV-LPM entails estimating an LPM model of LFP using SAH as an instrument (where the first stage is also LPM).

The outcome of interest is LFP, a dummy variable which equals one if the respondent is a labour force participant (broad definition, i.e. the respondent was willing to work in the past month) and zero otherwise. It is expected that a healthy individual will be more willing and able to work relative to one in poor health, ceteris paribus. Estimates from strict LFP models were used as a robustness check on the results.

For the full/total specification (i.e. consisting of both the male and female samples), X consists of household grant receipt (a proxy for household non-labour income), years of schooling, age dummies, location, race, provincial unemployment rate, marital status, gender, number of under-17 children in the household and household size. For the gender-specific models, X is identical to the above except for the exclusion of gender from both male and female equations and the inclusion of the presence of at least one employed male in the household in the female specification.

Regarding identification, we followed Sterns' (1989) and Cai and Kalb's (2006) strategy of instrumenting SAH with relatively objective health measures. The instruments are dummy variables indicating whether the respondent experienced joint pain/arthritis and/or memory loss in the past 30 days. This is hinged on the recognition that SAH is a summary/representative measure of overall health status, an assertion amply demonstrated in the literature (Benítez-Silva et al., 2004; Ferraro, 1980; LaRue, Bank, Jarvik, & Hetland, 1979; Nagi, 1969). We assume that an individual who experiences say, joint pains and memory loss would only change her labour market status due to illness only when such a condition so adversely affects her health that she rates it worse than "good". This informs the treatment of such relatively objective health conditions as SAH instruments given that much as they directly affect health status (proxied by SAH), there may be no compelling reason to directly include them in a LFP equation which includes SAH as a covariate. Thus, we assume that they affect LFP only indirectly through their influence on overall health status (captured by SAH). This is similar to Bound's (1991) argument that merely proxying health status with relatively objective health measures in a labour supply equation does not solve the problem of endogeneity as objective health measures do not perfectly reflect work capacity. But we are not unmindful of the fact that regarding SAH responses as representing the totality of true health status may come across as a strong assumption. Unfortunately, this is the much we can do in the present circumstances.

A potential argument against the suitability of these instruments is that unobserved individual characteristics associated with low LFP probability (e.g. low innate ability and drive) may also lead to a higher probability of having such health conditions (i.e. the instruments). To our best knowledge, there is no convincing evidence in the literature in this regard. Moreover, we regressed each of the instruments on parental education and parental mortality before the respondent turned five years (since these variables may affect a respondent's future labour market outcomes) as well as relevant covariates like own education and other socio-economic variables. Parental education and parental mortality were not statistically significant even at the 10% level for both males and females (results available on request).

Furthermore, though these instruments are self-reported, they are not likely to be the result of rationalization of labour market status (Bound, 1991; Bound et al., 1995; Cai, 2010; Cai & Kalb, 2006; Stern, 1989). This is important as their validity will be questionable if they are determined by one's employment status. For instance, if being employed under hazardous working conditions significantly determines any of the instruments (as may be obtainable for asthma among mine workers), or if unemployed respondents who do not actually suffer from any of the conditions implied by the instruments declare that they suffer from them due to the shame/stigma associated with being unemployed, the instruments may no longer be valid. To empirically test this, we regressed each instrument on the employment status dummy and a host of covariates like education, age, race and gender. The employment dummy was not statistically significant even at the 10% level in each of the regressions (results also available on request). This empirically strengthens the case for the exogeneity of the instruments. Finally, it is hoped that the array of socio-economic controls included in the health and LFP equations above helped purge the error terms of most of the plausible reasons why they may be conditionally correlated with the instruments.

Bound (1991) has observed that instrumenting SAH with more objective health variables may not be enough to identify the effect of health on LFP if SAH responses are informed by the expectation of financial rewards for disability (e.g. disability grants). In this case, if otherwise healthy but non-labour force participating individuals systematically report poor health as the reason for their non-participation (i.e. rationalization endogeneity) because they expect to receive disability benefits tied to non-participation, even our strategy of using relatively objective health measures as SAH instruments will yield inconsistent SAH coefficients. But this is not likely a significant issue in this study, as disability grant receipt in South Africa is not predicated on non-labour force participation. However, the estimates are consistent if rationalization endogeneity is due to, say, tastes for leisure (Au et al., 2005). Indeed, in results not reported here, the results of this study did not significantly change when disability grant recipients were excluded from the analysis.

This analysis was mainly based on the third (i.e. most recent) wave of the National Income Dynamics Study (NIDS) dataset while a sub-analysis was conducted with both waves 1 and 3 data. NIDS is a nationally representative panel dataset of South African households and a rich source of socio-economic data. Wave 3 data was collected in 2012 and each of the previous waves was collected two years apart. Detailed description of the dataset is available at www.nids.uct.ac.za. The sample was restricted to respondents aged 20-60 years in wave 3 so as to exclude students and retirees from the analysis. Furthermore, Asians were excluded given their small sample size.

RESULTS AND DISCUSSION

Table 1 shows descriptive statistics of variables used in the analysis.

[Table 1]

With regard to a priori expectations, education is expected to increase the probability of LFP. Also, it is likely to improve health through greater awareness of health-related knowledge (Cai & Kalb, 2006). Age is also likely to be positively correlated with LFP as older respondents tend to have more labour market-related networks. However, it is negatively associated with health status (Cai & Kalb, 2006; Kenkel, 1995). Location and race capture place and racial heterogeneity in health status. Compared to Africans and coloureds, whites are expected to have higher LFP owing to historical realities in the country. Also given that they are more likely to reside in urban formal centres with better medical facilities, whites are expected to enjoy better health status. Evidence of a significant relationship between marital status and LFP exists in the literature, where the relationship is especially negative for females (Jaumotte, 2003). However, health is mainly seen to be positively associated with marital status (Beckett & Elliott, 2002). Grant receipt is likely to be negatively associated with LFP as non-labour income may result in higher reservation wage or early retirement (Mastrobuoni, 2009). Provincial unemployment rate is expected to be negatively associated with LFP (Dinkelman & Pirouz, 2002; Evans & McCormick, 1994). This variable was used as a proxy for local labour market conditions, an attempt to capture some demand-side determinants of labour supply. Having an employed male (usually a primary breadwinner especially in an African context) may put less economic pressure on the woman, thereby resulting in reduced female LFP especially if the hypothesis that women are the main producers of domestic services like childbearing and child care holds (Joll, McKenna, McNabb, & Shorey, 1983); but it may also lead to a rise in female LFP due to increased labour market-related information/networks (Dinkelman & Pirouz, 2002).

Table 1 shows that 73% of the sample were labour force participants while 90% reported being in the excellent, very good or good health category. Married/cohabiting respondents made up 40% of the sample, while 48% were male. With regard to the SAH instruments, 8% and 5% suffered from joint pain/arthritis and memory loss respectively.

Figure 2 below reveals a positive relationship between better health status and LFP for both males and females. Apparently, there was not much heterogeneity in the LFP rates of the various groups classified as healthy. Furthermore, males had higher participation rates across all SAH categories (except poor) than females. Between-gender (i.e. male-female) percentage point difference ranged from -6 for the poor health category to 19 for the fair category. Figure 2 also highlights perhaps structural features of the South African labour market unrelated to health as broad LFP rate among men in fair health exceeded that of women in even excellent health.

[Figure 2]

Table 2 reveals that the proportion of non-labour force participants who self-reported being sick (i.e. fair or poor health) was twice that of labour force participants for both strict and broad LFP. This is a tentative indication of a positive relationship between better SAH and LFP.

[Table 2]

Also, we considered a descriptive relationship between SAH and its instruments in Table 3.

[Table 3]

Table 3 provides preliminary evidence of non-trivial correlation between SAH and the instruments. For instance among those with no joint pain/arthritis, only 8% reported being in fair or poor health while it was 40% for those with the condition. A similar situation obtained for memory loss.

Table 4 depicts marginal effects in both the bivariate probit and IV-LPM models. In line with Angrist and Pischke (2009), a useful way to think about policy-relevant measures in the bivariate probit framework is to consider marginal effects rather than difficult-to-interpret index coefficients.

[Table 4]

With the bivariate probit model, Angrist and Pischke have observed that one can consider the effect of an arguably endogenous regressor in a manner akin to the effect of a treatment on some outcome of interest. As noted above, one can recover estimates of ATE, TOT and LATE from the bivariate probit model. Here, being in "good" health may be considered as a kind of an endogenous treatment, while the instrument status is considered as an assignment to either treatment or control. Thus, ATE is the effect of the treatment on the outcome (LFP) assuming randomized treatment assignment. TOT is the effect of good health for those who reported being healthy irrespective of their instrument status. Finally, LATE is the effect of good health for those who reported being in good health only because they experienced memory loss and/or joint pains/arthritis, i.e. compliers; a sub-population that cannot be determined either a priori or ex post facto (Gennetian, Bos, & Morris, 2002). Admittedly, this may constitute a tiny proportion of the sample. Among these estimators, LATE is the only one possible with the IV-LPM (Angrist & Pischke, 2009). Similar measures have been recovered from both the bivariate probit and IV-LPM models in a study of the effect of fertility on female employment where similar identification issues arise as in this study (Angrist & Pischke, 2009). Therefore, following Angrist and Pischke, ATE, i.e.

$$E[l_1 - l_0] = E\{1[X/\alpha^* + \delta^* > \varepsilon_L] - 1[X/\alpha^* > \varepsilon_L]\}$$

can be specified as:

$$E\{1[X/\alpha^* + \delta^* > \varepsilon_L] - 1[X/\alpha^* > \varepsilon_L]\} = E\left\{\Phi\left[\frac{X/\alpha^* + \delta^*}{\sigma_{\varepsilon_L}}\right] - \Phi\left[\frac{X/\alpha^*}{\sigma_{\varepsilon_L}}\right]\right\} \qquad [7]$$

while TOT, i.e. $E[lfp_1 - lfp_0|SAH = 1]$ takes the following form:

$$E\{1[X/\alpha^* + \delta^* > \varepsilon_L] - 1[X/\alpha^* > \varepsilon_L]|X/\beta^* + \gamma^* z > \varepsilon_H\}$$
$$= E\left\{\frac{\Phi_b\left(\frac{X/\alpha^* + \delta^*}{\sigma_{\varepsilon_L}}, \frac{X/\beta^* + \gamma^* z}{\sigma_{\varepsilon_H}}; \rho_{\varepsilon_L\varepsilon_H}\right) - \Phi_b\left(\frac{X/\alpha^*}{\sigma_{\varepsilon_L}}, \frac{X/\beta^* + \gamma^* z}{\sigma_{\varepsilon_H}}; \rho_{\varepsilon_L\varepsilon_H}\right)}{\Phi\left(\frac{X/\beta^* + \gamma^* z}{\sigma_{\varepsilon_H}}\right)}\right\} \qquad [8]$$

Following Chiburis et al. (2011), the LATE estimator for the bivariate probit model is specified thus:

$$\frac{\left[\Phi_b\left(X/\beta^*+\gamma^*, X/\alpha^*+\delta^*; \rho_{\varepsilon_L\varepsilon_H}\right)+\Phi_b\left(-(X/\beta^*+\gamma^*), X/\alpha^*; -\rho_{\varepsilon_L\varepsilon_H}\right)\right]-\left[\Phi_b\left(X/\beta^*, X/\alpha^*+\delta^*; \rho_{\varepsilon_L\varepsilon_H}\right)+\Phi_b\left(-X/\beta^*, X/\alpha^*; -\rho_{\varepsilon_L\varepsilon_H}\right)\right]}{\Phi(X/\beta^*+\gamma^* z)-\Phi(X/\beta^*)} \qquad [9]$$

where l_0, l_1, Φ, σ and ρ are the LFP outcome of the non-treated, the LFP outcome of the treated, the normal CDF, error variance and error correlation coefficient respectively. Individual subscripts have been omitted for notational simplicity while the simultaneous estimation of the first and second stage equations is implicit in the above system. For IV-LPM, the LATE estimator is obtained as the usual marginal effect in a two stage least squares framework. Table 5 depicts ATE, LATE and TOT estimates of the impact of health on LFP obtained from the above bivariate probit and IV-LPM models.

[Table 5]

From Table 5, ATE was 0.23 for the total specification using equation [7]. For the female and male specifications, ATE was 0.20 and 0.29 respectively. Using equation [8], TOT estimates were 0.26, 0.23 and 0.33 for the total, female and male specifications respectively. As indicated above, the set of bivariate probit estimates that should be directly comparable with IV-LPM estimates are the LATE estimates (columns 7-9). Apart from the male coefficient, both models were quite numerically similar, with bivariate probit coefficients slightly higher than their IV-LPM counterparts [Abadie (2000b) and Angrist (2001) also found higher bivariate probit estimates relative to linear IV estimates]. Even the male bivariate probit LATE estimate that seemed to differ much from its IV-LPM counterpart was only statistically significant at 10%. Thus, Table 5 indicates that though the estimators capture the effect of health on LFP for different sub-populations as indicated in the above definitions, the effects were generally similar for each specification: 20-23%, 29-33%, and 23-26% among the female, male and total specifications respectively.

Linear SAH

There may be concerns with the “arbitrary” dichotomization of SAH as well as the very restrictive assumptions of the bivariate probit model. Consequently, the IV model above may be estimated using the original linear SAH index (recall that higher SAH values indicate worse self-reported health outcomes). The result (Table 6) indicates a nontrivial effect of a one unit deterioration in self-reported health on the probability of LFP especially for the total specification. As in the IV-LPM, the gender-based estimates were similar to the total estimates but statistically insignificant at conventional levels due to slightly higher standard errors. Thus, whether or not health is dichotomized, there is a significant impact of health on LFP. The controls were also similar to those in Table 4. Thus, the pattern of the results was similar to when SAH was dichotomized.

[Table 6]

First stage estimates and instrument relevance

Full first stage results for the IV-LPM are presented in Table A1 in the appendix (covariates conformed to theoretical expectations and were mostly statistically significant at conventional levels) while results pertaining only to the instruments are presented in Table 7. The results clearly show that each of the instruments significantly predicted SAH even at the 1% level of significance. They also conformed to theoretical expectations, as suffering from any of the conditions was associated with increased probability of being in poor health relative to not suffering from it. The F statistics in a joint test of the instruments in all specifications (see Table 7) exceeded the Staiger and Stock (1997) critical F statistic of 10, implying that the instruments were not weak. Similar conclusion was reached for the gender-based models.

[Table 7]

Instrument validity

To formally ascertain instrument validity, Hansen's J test of over-identifying restrictions failed to reject the null hypothesis of valid instruments even at 10% as shown in Table 7. With regard to whether relatively objective health conditions directly explained LFP, we included different relatively objective health measures (compared to SAH) in a LFP equation that excluded SAH[4]. The relatively objective health conditions include: serious injury, body ache, persistent cough, rash, painful urination, back ache and tight chest among others (results available on request). None of these other objective health controls was statistically significant even at the 10% level. This is expected and is in line with Bound et al. (1999) who maintained that these measures also suffer from measurement error in that they do not perfectly reflect work capacity (thus, making them susceptible to attenuation bias inherent in poorly-measured variables when used as health proxies in a labour supply regression).

A potential concern regards the generalizability of the above bivariate probit and IV-LPM results over the entire population as it may be argued that the instruments are likely to disproportionately affect old individuals. To ascertain whether the above results were driven by old respondents, we re-estimated the models on a sub-sample younger than 50 years. Across all estimators (ATE, TOT and LATE), the results were very similar to the above estimates, mainly differing in the second decimal place. Also, the patterns of statistical significance were similar. These results are available on request. The same is true for the IV-LPM results.

With regard to the controls in Table 4 above, the marginal effect of interest for the kth continuous covariate in the LFP equation is $\frac{\partial \Pr(lfp=1|X_{-k})}{\partial X_k}$ (where X_{-k} are covariates other than X_k). This is the marginal effect of each covariate on the probability of a respondent being a labour force participant conditional on other covariates. For each discrete covariate, the change in LFP probability is in response to a unit change in such a covariate.

Average marginal effects of regression controls from both specifications (i.e. bivariate probit and IV-LPM) were almost identical and the estimates largely conformed to a priori expectations and were statistically significant at conventional levels. For instance, household grant receipt was associated with 5-7% reduction in LFP probability across both models, a

[4] The bias caused by including the potentially endogenous SAH in this regression has been noted by Murray (2006).

finding similar to Bertrand et al.'s (2003) estimate of 4% reduction of employment probability associated with household eligibility for the old age pension among Africans. Also, having at least a matric (i.e. $\geq$ 12 years of schooling) was associated with 5-15% increase in LFP probability. Furthermore, spatial distribution mattered for LFP (apparently reflecting persistent effects of apartheid-era living arrangements, where most jobs were located away from informal areas mostly occupied by non-whites); LFP probability was generally higher in other locations relative to traditional authority areas mainly populated by Africans. Also, the age-LFP gradient increased up to the mid-forties for men, after which it declined (albeit still positive and statistically significant). For women, it steadily increased up to the late thirties and declined afterwards. This finding is similar to Ntuli and Wittenberg (2013), where only the oldest age cohort (in their case, aged 55-59 years) did not have significantly higher LFP relative to 20-24 year old African women. Marriage/cohabitation was negatively (positively) associated with female (male) participation, thereby supporting the theory of marriage and labour market participation which suggests that single women are more likely to be economically active compared to their married counterparts while the converse holds for men (Becker, 1981). Negative marital status-LFP relationship for women has also been documented in South Africa (Ntuli & Wittenberg, 2013; Posel & Casale, 2003). Males had 10% higher LFP rates than females. For females, the positive sign on the presence of at least one employed male in the household negates the added worker effect (Jaumotte, 2003), suggesting that male participation was associated with increased female participation. This likely captures network effects as suggested by Dinkelman and Pirouz (2002). The flip side is that families without male participants may become poorer as not having at least an employed male in the household was associated with 4-5% decline in the probability of female participation.

Results in Table 4 (for both SAH and controls) were robust to the outcome being strict LFP except that the female SAH coefficient was no longer statistically significant at conventional levels. These results are available on request.

Testing for exogeneity of SAH

Though global self-reported health has been argued to be a plausibly endogenous determinant of LFP theoretically, some studies could not reject the hypothesis of exogenous SAH. If SAH is truly an exogenous determinant of LFP in any empirical context, applying IV techniques may be superfluous and will increase the likelihood erroneously finding no impact given the relative inefficiency of IV estimates compared to ordinary least squares (OLS) (Cameron & Trivedi, 2010; Murray, 2006).

A formal test of SAH exogeneity in the bivariate probit framework is a Wald test of the statistical significance of the cross-equation correlation (ρ) between the LFP and SAH equations. If ρ is statistically different from zero, there is evidence of SAH endogeneity and vice versa (Cai & Kalb, 2006; Stern, 1989). Table 4 above suggests failure to reject the null hypothesis of SAH exogeneity, i.e. $\rho = 0$, across all specifications. Also, Durbin-Wu-Hausman tests of the null hypothesis that SAH is exogenous failed to reject the null hypothesis of SAH exogeneity in all three specifications given p values of 0.24, 0.54 and 0.72 for the total, female and male specifications respectively. These suggest that SAH may not be endogenous to LFP in this context. The above conclusion of no SAH endogeneity is not novel as Stern (1989) found evidence of weakly endogenous self-reported disability on LFP in the US. Indeed, Bound et al. (1999) observed that most of the literature on the effect of health on labour force behaviour treats health as exogenous. On the contrary, some studies have rejected the hypothesis of exogeneity of SAH in Australia (Cai, 2010; Cai & Kalb, 2006). Thus, while SAH might prove to be exogenous in certain contexts, such conclusions appear to be context-specific and should be informed by appropriate statistical tests.

Marginal effects from a regression of LFP on SAH under the assumption of SAH exogeneity are shown in Table 8 below. Comparing these estimates for the regression controls with those in Table 4, one observes that the models appear to be robust to the exogeneity assumption. Perhaps this reinforces the findings from the more formal Durbin-Wu-Hausman and error correlation tests above. However, there was a marked decline in the marginal effect of SAH especially in the probit model in Table 8 (relative to the bivariate probit marginal effects). Here, being healthy was associated with 10-11%, 12-14% and 10-12% increase in LFP probability in the female, male and general populations respectively. However, the relative efficiency of single equation estimates relative to IV estimates resulted in statistically significant SAH marginal effects across all specifications even with these numerically smaller estimates.

[Table 8]

Past health versus current LFP

Another way to look at the relationship between health and LFP if one suspects endogeneity of SAH is to exploit the timing of SAH responses relative to LFP status. Given the panel nature of the dataset, one can estimate current LFP as a function of past SAH, an approach that has been adopted in the literature (Bartel & Taubman, 1986; Haveman et al., 1994). Though this method does not completely solve the endogeneity problem especially if past SAH responses were used to justify past LFP status, it does mitigate the contemporaneous feedback relationship that likely exists between both variables. Therefore, we regressed wave 3 LFP status on wave 1 SAH status and the same controls used above including wave 1 LFP. The marginal effects of past SAH are presented in Table 9 below (marginal effects of the controls are similar to what obtained in the main regressions and are therefore not reported; they are however available on request).The results (expectedly attenuated) show that health significantly affected LFP even four years after a given health assessment. The magnitude ranged from 6-9%.

[Table 9]

Health shocks and labour force participation

Some studies have exploited health changes (otherwise termed health shocks) available in panel datasets to provide a plausible argument with regard to the labour supply effect of arguably exogenous variations in SAH (Chirikos & Nestel, 1985; Gómez & Nicolás, 2006). Consequently, we estimated models for the effect of positive and negative self-reported health shocks on LFP. We categorized an individual as having suffered an adverse health shock if she identified herself as healthy in wave 1 but sick in wave 3 while the baseline category consisted of respondents who were healthy in both waves. On the other hand, an individual was identified as having experienced health improvement if she was sick in wave 1 but healthy in wave 3 while the benchmark category in this case comprised respondents who remained sick in both waves. For the first category, 7.6% of usable responses (total=7112 observations) recorded adverse health shocks among working age respondents in wave 3 while 71.9% of usable responses from the second category (total=1333 observations) recorded health improvement.

As Table 10 indicates, the results for adverse health shocks were qualitatively similar (in terms of sign) to what obtained in the contemporaneous and between-wave analyses above: adverse health shocks were associated with 12-14% decline in male LFP in wave 3 while such shocks were not associated with any change in female LFP. These estimates were however smaller than the IV estimates in Table 4 and Table 5. The "total" coefficients representing 8-9% higher probability of inactivity due to adverse health shocks in the population is higher than Gomez and Nicholas' (2006) estimate of 3-4% in Spain. Such differences may be due to the fact that Gomez and Nicholas estimated causal effects of adverse health shocks (which this study could not recover due to fewer data waves). They could also stem from apparent differences in levels of mechanization in both labour markets as South Africa is a developing economy where active labour market participation may be associated with a steeper gradient with health. Health improvement was statistically significant for both males and females. Labour force participation probability for those who experienced health improvement increased by 10-11%, 14-16% and 11% in the female, male and general populations respectively relative to those who remained in constant bad health. Thus, both health deterioration and health improvement were associated with higher percentage changes in LFP probability for men relative to women. This is consistent with all the evidence throughout this paper.

[Table 10]

CONCLUSION

This paper has ascertained the effect of health on labour force participation in South Africa. The foregoing discussion suggests that better health (measured as composite/global self-reported health) is positively related to LFP irrespective of the estimator. When treated as endogenous in a contemporaneous setting, the results suggest that SAH exerts a positive effect on LFP for the full/total, female and male samples. For the ATE, we find a 20-29% effect of health across specifications. The effect of SAH on LFP for those who self-reported "good" health irrespective of their memory loss and joint pains/arthritis status (i.e. TOT) is slightly higher than ATE: 23-33% across specifications. For those individuals who reported "good" health only because they did not suffer from memory loss or joint pains/arthritis (i.e. compliers), the effects (i.e. LATE) are similar to the ATE: 20-29% effect across the different specifications. Using the linear SAH index, a one point health deterioration results in an 8% decline in LFP probability. But various diagnostic tests suggest that SAH may not be endogenous to LFP in South Africa. Exploring dynamic health changes, adverse health shocks are associated with significantly decreased LFP of around 12-14% and 8-9% in both the male and total specifications respectively over the four-year period (2008-2012). Health improvement is associated with about 11% LFP increase among the female and total specifications as well as 16% increase among males. This study has therefore shown that better health is positively related to labour force participation in South Africa and that the relationship is not just temporary.

REFERENCES

Abadie, A. (2000). Semiparametric estimation of instrumental variable models for causal effects. *Technical Working Paper 260, National Bureau of Economic Research, Cambridge, MA.*

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business & Economic Statistics, 19*(1), 1-28.

Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, New Jersey: Princeton University Press.

Ardington, C., & Gasealahwe, B. (2014). Mortality in South Africa: Socio-economic profile and association with self-reported health. *Development Southern Africa, 31*(1), 127-145.

Arndt, C., & Lewis, J. D. (2001). The HIV/AIDS pandemic in South Africa: Sectoral impacts and unemployment. *Journal of International Development, 13*(4), 427-449.

Au, D., Crossley, T. F., & Schellhorn, M. (2005). The effect of health changes and long-term health on the work activity of older canadians. *Health Economics, 14*(10), 999-1018.

AVERT (undated). South Africa HIV/AIDS statistics. http://www.avert.org/south-Africa-hiv-aids-statistics.htm. Accessed 30 May 2015.

Banerjee, A., Galiani, S. F., Levinsohn, J. A., McLaren, Z., & Woolard, I. (2008). Why has unemployment risen in the new South Africa. *Economics of Transition, 16*(4), 715-740.

Bartel, A., & Taubman, P. (1986). Some economic and demographic consequences of mental illness. *Journal of Labor Economics, 4*, 243-256.

Becker, G. S. (1981). *A treatise on the family* Cambridge, Harvard University Press.

Beckett, M., & Elliott, M. N. (2002). Does the association between marital status and health vary by sex, race, and ethnicity? *Rand Working Paper Series 02-08.*

Benítez-Silva, H., Buchinsky, M., Chan, H. M., Cheidvasser, S., & Rust, J. (2004). How large is the bias in self-reported disability? *J.Appl.Econ, 19*, 649-670.

Bertrand, M., Mullainathan, S., & Miller, D. (2003). Public policy and extended families: Evidence from pensions in South Africa. *The World Bank Economic Review, 17*(1), 27-50.

Bhorat, H. (2007). Unemployment in South Africa: Descriptors and determinants. *Unpublished Consultant Report for OECD, Paris.*

Booysen, F. R., Bachmann, M., Matebesi, Z., & Meyer, J. (2002). The socioeconomic impact of HIV/AIDS on households in South Africa: Pilot study in Welkom and QwaQwa, Free State province. *University of the Free State.*

Boskin, M. J. (1977). Social security and retirement decisions. *Economic Inquiry, 15*(1), 1-25.

Bound, J. (1989). The health and earnings of rejected disability insurance applicants. *NBER Working Paper no. 2816.*

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources,* (1), 138.

Bound, J., Schoenbaum, M., Stinebrickner, T., & Waidmann, T. (1999). The dynamic effects of health on the labor force transitions of older workers. *Labour Economics, 6*(2), 179-202.

Bound, J., Schoenbaum, M., & Waidmann, T. (1995). Race and education differences in disability status and labour force attachment in the health and retirement survey. *Journal of Human Resources, 30,* S227-S267.

Bradshaw, D., Groenewald, P., Laubscher, R., Nannan, N., Nojilana, B., Norman, R., . . . Schneider, M. (2000). *Initial burden of disease estimates for South Africa.* (). Cape Town: South African Medical Research Council.

Bridges, S., & Lawson, D. (2008). *Health and labour market participation in Uganda* United Nations University. World institute for development economics research (UNU/WIDER).

Cai, L. (2010). The relationship between health and labour force participation: Evidence from a panel data simultaneous equation model. *Labour Economics, 17*(1), 77-90.

Cai, L., & Kalb, G. (2006). Health status and labour force participation: Evidence from Australia. *Health Economics, 15*(3), 241-261.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. New York: Cambridge University Press.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata* Stata Press College Station, TX.

Campolieti, M. (2002). Disability and the labor force participation of older men in Canada. *Labour Economics, 9*(3), 405-432.

Chiburis, R. C., Das, J., & Lokshin, M. (2011). A practical comparison of the bivariate probit and linear IV estimators. *Policy Research Working Paper 5601.*

Chirikos, T. N., & Nestel, G. (1985). Further evidence on the economic effects of poor health. *The Review of Economics and Statistics,* 61-69.

Choi, H. (2007). *Health impacts on future labor market outcome: Exploring PSID siblings.* (Unpublished Citeseer).

Currie, J., & Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of Labor Economics, 3,* 3309-3416.

Dinkelman, T., & Pirouz, F. (2002). Individual, household and regional determinants of labour force attachment in South Africa. *South African Journal of Economics, 70*(5), 865-891.

Dwyer, D. S., & Mitchell, O. S. (1999). Health problems as determinants of retirement: Are self-rated measures endogenous? *Journal of Health Economics, 18*(2), 173-193.

Evans, P., & McCormick, B. (1994). The new pattern of regional unemployment: Causes and policy significance. *The Economic Journal, 104*(424), 633-647.

Ferraro, K. F. (1980). Self-ratings of health among the old and the old-old. *Journal of Health and Social Behavior,* 377-383.

Gennetian, L. A., Bos, J. M., & Morris, P. A. (2002). Using instrumental variables analysis to learn more from social policy experiments. MDRC working papers on research methodology. *MDRC Working Papers on Research Methodology.*

Godlonton, S., & Keswell, M. (2005). The impact of health on poverty: Evidence from the South African integrated family survey. *South African Journal of Economics, 73*(1), 133-148.

Gómez, P. G., & Nicolás, A. L. (2006). Health shocks, employment and income in the spanish labour market. *Health Economics, 15*(9), 997-1009.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society, 45*, 1-22.

Haan, P., & Myck, M. (2009). *Dynamics of poor health and non-employment* Forschungsinst. zur Zukunft der Arbeit.

Haveman, R., Wolfe, B., Kreider, B., & Stone, M. (1994). Market work, wages, and men's health. *Journal of Health Economics, 13*(2), 163-182.

Jack, W., & Lewis, M. (2009). Health investments and economic growth: Macroeconomic evidence and microeconomic foundations. *World Bank Policy Research Working Paper no.4877.*

Jaumotte, F. (2003). Female labour force participation: Past trends and main determinants in OECD countries. *OECD Economics Department Working Papers, no. 376.*

Joll, C., McKenna, C. J., McNabb, R., & Shorey, J. (1983). *Developments in labour market analysis* G. Allen & Unwin.

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science & Medicine, 69*(3), 307-316.

Kenkel, D. S. (1995). Should you eat breakfast? Estimates from health production functions. *Health Economics, 4*(1), 15-29.

Kreider, B. (1999). Latent work disability and reporting bias. *Journal of Human Resources,* 734-769.

LaRue, A., Bank, L., Jarvik, U., & Hetland, M. (1979). Health in old age: How do physicians' ratings and self-ratings compare? *Journal of Gerontology, 34*(5), 687-691.

Levinsohn, J., McLaren, Z. M., Shisana, O., & Zuma, K. (2013). HIV status and labor market participation in South Africa. *The Review of Economics and Statistics, 95*(1), 98-108.

Manning, W. G., Newhouse, J. P., & Ware Jr, J. E. (1982). The status of health in demand estimation; or, beyond excellent, good, fair, poor. In V. R. Fuchs (Ed.), *Economic aspects of health* (pp. 143-181). Chicago, Illinois: University of Chicago Press.

Mastrobuoni, G. (2009). Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities. *Journal of Public Economics, 93*(11), 1224-1233.

Mete, C., & Schultz, T. P. (2002). Health and labor force participation of the elderly in Taiwan. *Economic Growth Center Discussion Paper, no. 846, Yale University*.

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *The Journal of Economic Perspectives, 20*(4), 111-132.

Nagi, S. Z. (1969). Congruency in medical and self-assessment of disability. *IMS, Industrial Medicine and Surgery, 38*(3), 27-36.

Novignon, J., Nonvignon, J., & Arthur, E. (2015). Health status and labour force participation in Sub-Saharan Africa: A dynamic panel data analysis. *African Development Review, 27*(1), 14-26.

Ntuli, M., & Wittenberg, M. (2013). Determinants of black women's labour force participation in post-apartheid South Africa. *Journal of African Economies, 0*(0), 1-28.

O'Donnell, O. (1998). The effect of disability on employment allowing for work incapacity. *University of Kent at Canterbury Department of Economics Working Paper Series, 46*(98-13).

Parsons, D. O. (1980). The decline in male labor force participation. *The Journal of Political Economy*, *88*(1), 117-134.

Posel, D., & Casale, D. (2003). What has been happening to internal labour migration in South Africa, 1993–1999? *South African Journal of Economics*, *71*(3), 455-479.

Smith, J. P. (2009). The impact of childhood health on adult labor market outcomes. *The Review of Economics and Statistics, 91*(3), 478-489.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557-586.

Statistics South Africa. (2012). *Census 2011 statistical release.* (No. P0301.4). Pretoria: Statistics South Africa.

Stern, S. (1989). Measuring the effect of disability on labor force participation. *Journal of Human Resources,* 361-395.

WHO. (2009). *Mortality and burden of disease estimates for WHO member states in 2004.*

Wittenberg, M. (1999). A spatial analysis of unemployment. *Symposium on Spatial Guidelines for Infrastructure, Investment and Development. Office of the Deputy President. Pretoria.*

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Young, A. (2005). The gift of the dying: The tragedy of AIDS and the wealth of future African generations. *Quarterly Journal of Economics, 120*(2), 423-466.

APPENDIX

Table A1: First stage estimates: IV-LPM

	(1)	(2)	(3)
Variables	Total	Female	Male
joint pain	-0.25***	-0.25***	-0.23***
	(0.03)	(0.03)	(0.06)
memory loss	-0.17***	-0.15***	-0.20***
	(0.04)	(0.04)	(0.06)
grant	-0.02***	-0.02*	-0.02
	(0.01)	(0.01)	(0.01)
matric	0.03***	0.04***	0.02*
	(0.01)	(0.01)	(0.01)
age26-30	-0.02*	-0.02	-0.01
	(0.01)	(0.02)	(0.01)
age31-35	-0.06***	-0.07***	-0.05***
	(0.02)	(0.03)	(0.02)
age36-40	-0.05***	-0.06***	-0.03**
	(0.01)	(0.02)	(0.01)
age41-45	-0.08***	-0.07***	-0.11***
	(0.02)	(0.02)	(0.03)
age46-50	-0.11***	-0.11***	-0.12***
	(0.02)	(0.03)	(0.02)
age51-60	-0.14***	-0.16***	-0.11***
	(0.02)	(0.03)	(0.02)
rural formal	-0.06**	-0.08**	-0.02
	(0.02)	(0.03)	(0.03)
urban formal	-0.01	-0.03*	0.01
	(0.01)	(0.02)	(0.01)
urban informal	-0.02*	-0.04*	-0.00
	(0.01)	(0.02)	(0.02)
african	-0.02	-0.03	-0.01
	(0.02)	(0.03)	(0.03)
coloured	-0.03	-0.03	-0.03
	(0.02)	(0.03)	(0.03)
prov. unemp†	-0.00**	-0.00	-0.01**
	(0.00)	(0.00)	(0.00)
married††	0.04***	0.05***	0.02
	(0.01)	(0.01)	(0.01)
male	0.02***		
	(0.01)		
num. children‡	0.00	-0.00	0.00
	(0.00)	(0.01)	(0.01)
household size	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)
employed male‡‡		-0.01	
		(0.01)	
constant	1.08***	1.05***	1.13***
	(0.04)	(0.05)	(0.06)
R^2 adjusted	0.13	0.14	0.10
N	9795	5836	3956
F-stat	27.2	20.8	7.6

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition; †provincial unemployment rate; ††married/cohabiting; ‡number of under-17 children in household; ‡‡household has at least one employed male

Table 1: Descriptive statistics

Variable	N	Mean	Std.Dev.[a]
labour force participation	11250	0.73	0.4
self-assessed health	9822	0.90	0.3
matric (i.e. ≥ 12 years of schooling)	11271	0.37	0.5
age20 - 25	11294	0.22	0.4
age26 - 30	11294	0.15	0.4
age31 - 35	11294	0.14	0.3
age36 - 40	11294	0.13	0.3
age41 - 45	11294	0.11	0.3
age46 - 50	11294	0.10	0.3
age51 - 60	11294	0.15	0.4
traditional authority	11264	0.29	0.5
rural formal	11264	0.06	0.2
urban formal	11264	0.53	0.5
urban informal	11264	0.12	0.3
african	11294	0.83	0.4
coloured	11294	0.09	0.3
white	11294	0.08	0.3
married/cohabiting	9824	0.40	0.5
number of under-17 children in household	11294	1.43	1.7
male	11264	0.48	0.5
grant	11264	0.54	0.5
household has employed male	11285	0.50	0.5
provincial unemployment rate	11294	25.47	3.5
household size	11264	4.80	3.2
joint pain	9823	0.08	0.3
memory loss	9819	0.05	0.2

Source: NIDS wave 3; author's calculations; sample corrected for survey design and non-random attrition; [a] standard deviation

Table 2: Health-LFP status in wave 3 (row percentages sum to 100)

	Health Status (%)		Row total (N)
LFP Status	Sick	Healthy	
Non-LFP	16.0	84.0	**3217**
LFP	8.0	92.0	**6585**

Source: NIDS; author's calculations; sample corrected for survey design and non-random attrition

Table 3: Distribution of instruments across SAH categories

		SAH categories (%)		Row size (N)
	Has condition?	Sick	Healthy	
joint pain	No	7.7	92.3	**8949**
	Yes	40.2	59.8	**870**
Memory loss	No	9.0	91.0	**9295**
	Yes	35.0	65.0	**520**

Source: NIDS wave 3; author's calculations; sample corrected for survey design and non-random attrition

Table 4: Marginal effects of SAH and controls on LFP probability (bivariate probit and IV-LPM)

	(1)	(2)	(3)	(4)	(5)	(6)
		Bivariate probit			IV-LPM	
Variables	Total	Female	Male	Total	Female	Male
sah	0.22***	0.20*	0.25*	0.22**	0.17	0.19
	(0.07)	(0.11)	(0.14)	(0.08)	(0.11)	(0.13)
grant	-0.06***	-0.05***	-0.07***	-0.07***	-0.06***	-0.07***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
matric[a]	0.10***	0.15***	0.05*	0.10***	0.15***	0.05*
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
age26-30	0.14***	0.18***	0.09***	0.15***	0.19***	0.11***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
age31-35	0.15***	0.18***	0.12***	0.16***	0.19***	0.12***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
age36-40	0.19***	0.21***	0.15***	0.19***	0.22***	0.14***
	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)
age41-45	0.15***	0.15***	0.15***	0.16***	0.16***	0.14***
	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)	(0.04)
age46-50	0.11***	0.13***	0.05	0.11***	0.13***	0.06
	(0.02)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
age51-60	0.01	0.02	-0.03	-0.00	-0.00	-0.03
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
rural formal	0.07***	0.03	0.11***	0.09***	0.04	0.12***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
urban formal	0.11***	0.12***	0.08***	0.12***	0.13***	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
urban informal	0.06**	0.09***	0.02	0.08**	0.11***	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
african	0.03	0.05	-0.01	0.03	0.05	0.01
	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)
coloured	0.00	0.06	-0.08	0.01	0.06	-0.06
	(0.05)	(0.06)	(0.06)	(0.04)	(0.05)	(0.06)
prov. unemp[†]	0.00	-0.00	0.00*	0.00	-0.00	0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
married[††]	0.02	-0.04*	0.09***	0.02	-0.04*	0.08***
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
male	0.10***			0.10***		
	(0.01)			(0.01)		
num. children[‡]	-0.00	-0.00	-0.00	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
household size	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
employed male[‡‡]		0.04**			0.05**	
		(0.02)			(0.02)	
constant				0.26**	0.31**	0.40**
				(0.12)	(0.15)	(0.17)
R^2				0.12	0.12	0.11
F-stat	42.14	26.65	13.36	47.8	27.3	17.3
rho	-0.23	-0.20	-0.28			
	(0.14)	(0.19)	(0.30)			
N	9775	5825	3950	9775	5825	3950

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition; [a] ≥ 12 years of schooling; [†]provincial unemployment rate; [††]married/cohabiting; [‡]number of under-17 children in household; [‡‡]household has at least one employed male

Table 5: Effect of health on LFP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables	Bivariate probit									IV-LPM marginal effects		
	ATE			TOT			LATE			LATE		
	Total	Female	Male	Total	Female	Male	Total	Female	Male	Total	Female	Male
sah	0.23***	0.20**	0.29*	0.26***	0.23**	0.33*	0.24***	0.20**	0.29*	0.22**	0.17	0.19
	(0.08)	(0.09)	(0.17)	(0.09)	(0.10)	(0.09)	(0.08)	(0.08)	(0.17)	(0.08)	(0.11)	(0.13)
N	9775	5825	3950	9775	5825	3950	9775	5825	3950	9775	5825	3950

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition; standard errors obtained through 400 Monte Carlo replications

Table 6: Impact of health (linear SAH) on LFP

	(1)	**(2)**	**(3)**
Variables	Total	Female	Male
linear sah	-0.08***	-0.06	-0.07
	(0.03)	(0.04)	(0.04)
grant	-0.06***	-0.05**	-0.07**
	(0.02)	(0.02)	(0.03)
matric[a]	0.09***	0.15***	0.04
	(0.02)	(0.02)	(0.03)
age26-30	0.15***	0.20***	0.11***
	(0.02)	(0.03)	(0.03)
age31-35	0.16***	0.19***	0.13***
	(0.02)	(0.03)	(0.03)
age36-40	0.20***	0.23***	0.14***
	(0.02)	(0.03)	(0.03)
age41-45	0.17***	0.17***	0.15***
	(0.03)	(0.03)	(0.05)
age46-50	0.13***	0.15***	0.08
	(0.03)	(0.04)	(0.05)
age51-60	0.02	0.01	-0.02
	(0.03)	(0.04)	(0.05)
rural formal	0.08***	0.03	0.11***
	(0.03)	(0.04)	(0.04)
urban formal	0.11***	0.12***	0.08***
	(0.02)	(0.03)	(0.02)
urban informal	0.07**	0.10***	0.03
	(0.03)	(0.03)	(0.04)
african	0.02	0.05	-0.01
	(0.04)	(0.05)	(0.04)
coloured	0.00	0.06	-0.07
	(0.04)	(0.05)	(0.05)
prov. unemp[†]	0.00	-0.00	0.00*
	(0.00)	(0.00)	(0.00)
married[††]	0.02	-0.04*	0.08***
	(0.01)	(0.02)	(0.02)
male	0.10***		
	(0.01)		
num. children[‡]	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)
household size	-0.00	-0.01	-0.01
	(0.00)	(0.01)	(0.01)
employed male[‡‡]		0.05**	
		(0.02)	
constant	0.61***	0.58***	0.70***
	(0.08)	(0.11)	(0.11)
R^2	0.10	0.10	0.09
F-stat	48.72	26.63	17.40
N	9775	5825	3950

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition; [a] ≥ 12 years of schooling; [†]provincial unemployment rate; [††]married/cohabiting; [‡]number of under-17 children in household; [‡‡]household has at least one employed male

Table 7: First stage LPM estimates: marginal effects

	Dependent variable: $\mathbf{Pr}(\boldsymbol{sah} = \mathbf{1}\vert\boldsymbol{X})$		
	(1)	(2)	(3)
Variables	Total	Female	Male
joint pain	-0.25***	-0.25***	-0.23***
	(0.03)	(0.03)	(0.06)
memory loss	-0.17***	-0.15***	-0.20***
	(0.04)	(0.04)	(0.06)
N	9795	5836	3956
F (for joint pain & memory loss)	60.5	56.9	16.5
Hansen J (p value)	0.11	0.51	0.50

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition

Table 8: LFP determination allowing for exogeneity of SAH (marginal effects)

	(1)	(2)	(3)	(4)	(5)	(6)
		Probit			LPM	
Variables	Total	Female	Male	Total	Female	Male
sah	0.10***	0.10***	0.12***	0.12***	0.11***	0.14***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)
grant	-0.07***	-0.06***	-0.08***	-0.07***	-0.06***	-0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
matric	0.11***	0.16***	0.05**	0.11***	0.16***	0.05**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
age26_30	0.14***	0.18***	0.09***	0.15***	0.19***	0.11***
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
age31_35	0.15***	0.17***	0.11***	0.15***	0.18***	0.12***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
age36_40	0.19***	0.21***	0.15***	0.18***	0.22***	0.13***
	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)
age41_45	0.14***	0.14***	0.13***	0.15***	0.15***	0.13***
	(0.02)	(0.03)	(0.05)	(0.02)	(0.03)	(0.04)
age46_50	0.09***	0.12***	0.04	0.10***	0.12***	0.06
	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
age51_60	-0.01	-0.00	-0.05*	-0.02	-0.02	-0.04
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
rural formal	0.07***	0.02	0.11***	0.08***	0.03	0.12***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
urban formal	0.11***	0.11***	0.08***	0.12***	0.12***	0.09***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
urban informal	0.06**	0.09***	0.02	0.08**	0.10***	0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
african	0.03	0.04	-0.01	0.03	0.05	0.01
	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)
coloured	-0.00	0.06	-0.08	0.01	0.06	-0.06
	(0.05)	(0.06)	(0.06)	(0.04)	(0.05)	(0.05)
prov. unemp†	0.00	-0.00	0.00*	0.00	-0.00	0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
married††	0.02	-0.03	0.10***	0.02	-0.03*	0.08***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
male	0.11***			0.11***		
	(0.01)			(0.01)		
num. children‡	-0.00	-0.00	-0.00	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
household size	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
employed male‡‡		0.04**			0.05**	
		(0.02)			(0.02)	
constant				0.36***	0.38***	0.44***
				(0.08)	(0.10)	(0.10)
R^2				0.13	0.12	0.11
F-Stat	36.3	20.1	12.9	50.0	28.0	18.1
N	9784	5830	3954	9784	5830	3954

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition; †provincial unemployment rate; ††married/cohabiting; ‡number of under-17 children in household; ‡‡household has at least one employed male

Table 9: Relationship between past SAH and current LFP

	(1)	(2)	(3)	(4)	(5)	(6)
		Probit			LPM	
Variables	Total	Female	Male	Total	Female	Male
sah	0.06***	0.07**	0.09**	0.07***	0.07**	0.09**
	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)
F statistic	38.4	29.8	17.4	56.4	29.8	17.4
N	8328	5217	3111	8328	5217	3111

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition

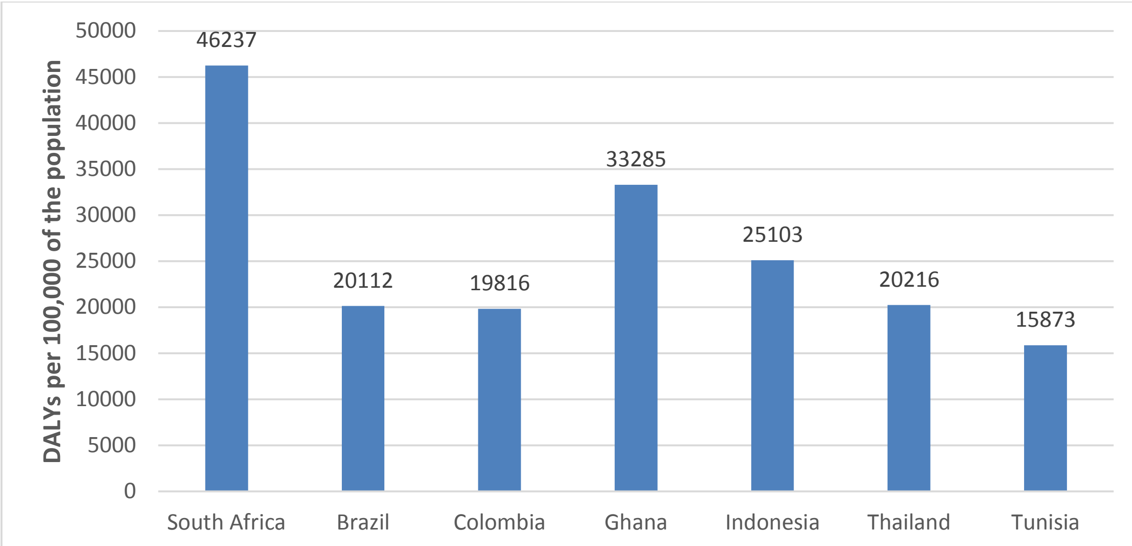
Table 10: Health shocks and LFP (marginal effects)

	(1)	(2)	(3)	(4)	(5)	(6)
		Probit			LPM	
Variables	Total	Female	Male	Total	Female	Male
			Adverse health shock			
	-0.08***	-0.05	-0.12***	-0.09***	-0.06	-0.14***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)
F statistic	26.5	14.1	11.4	34.6	18.1	14.1
N	7083	4318	2765	7083	4318	2765
			Health improvement			
	0.11***	0.10**	0.16***	0.11***	0.11**	0.14**
	(0.04)	(0.05)	(0.06)	(0.04)	(0.05)	(0.06)
F statistic	7.2	5.9	5.2	11.3	9.8	8.0
N	1327	947	380	1327	947	380

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; estimates corrected for survey design and non-random attrition

IMPACT OF HEALTH ON LABOUR FORCE PARTICIPATION IN SOUTH AFRICA

Figure 1: Absolute disease burden across select developing countries, 2004



Source: ECONEX calculations from WHO (2009)

Figure 2: Health-LFP relationship by gender