UNIVERSITY OF CAPE TOWN

SOCIO-ECONOMIC STATUS PROXIES IN STUDIES OF FERTILITY IN DEVELOPING COUNTRIES:

MEASURES AND METHODS

Leesen Naidu

Masters of Business Science - Economics

Mini-Dissertation, February 2009

Supervisor: Murray Leibbrandt

ABSTRACT

This paper explores the different measures and methods used to integrate socio-economic status into fertility analysis. The most common current practice is to create a single proxy summary measure and enter it as an independent variable in a multiple regression. I compare the consequences of using the different proxies on the estimated impact of socio-economic status and other determinants of fertility. I also introduce the use of a superior method of capturing the effect of socio-economic status on childbearing. Using data from a uniquely designed micro survey of young adults from Cape Town, South Africa, I find that a proxy based on household expenditure is the best predictor of fertility. Furthermore, the choice of proxy substantially affects the estimated impact of certain control variables. I finally demonstrate that the reduction in attenuation bias resulting from employing the new procedure can be significant. In addition, I find the use of an extended list of asset and housing quality variables most beneficial in certain circumstances. My findings also emphasise the sensitivity of the microanalyses to the treatment of SES that extends beyond that of fertility studies.

ACKNOWLEDGEMENTS

I would like to express my appreciation and thanks to my supervisor, Prof. Murray Leibbrandt, for his guidance, support and patience throughout the project.

I would also like to thank Martin Wittenberg for his critical input and assistance.

Final thanks must go out to Melina for her unmatched support and encouragement.

1 – INTRODUCTION

Socio-economic status (SES) is deeply rooted in fertility analysis simply because raising children requires a continuous flow of economic resources. Almost all studies include some measures of SES as either having direct influences on fertility or being used to control for unobserved factors. Regardless of the diverse literature on the causes and correlations of childbearing, there is a consensus that SES must be accounted for (Easterlin, 1969; Robinson, 1997). This extensive use of SES is, however, not accompanied by consistent definitions, measures or methods, thus leading to diverging results (Bollen, Glanville, & Stecklov, 2001). This inconsistency hinders our ability to analyse, not only the effect of SES on fertility, but also the effects of other determinants after controlling for SES.

The primary purpose of this paper is to investigate the consequences of different ways of measuring SES on its own estimated impact and the impact of other determinants of fertility. In this regard, the objective is not to identify what the actual effects on fertility are, but rather how the effects differ depending on the choice of SES proxy. I follow Bollen, Glanville, & Stecklov (2002), using a basic fertility model and focus on the dominant practice of using a *single* proxy variable. I also introduce a new method of integrating *multiple* SES proxies into fertility analysis that is based on the Lubotsky-Wittenberg estimation procedure as discussed in Lubotsky and Wittenberg (2006).

Recent research reflects a growing interest in the issue of the measurement of SES in studies of fertility and health in developing countries. Montgomery et al. (2000) outline a number of current practices that attempt to measure living standards. Filmer and Pritchett (2001) employ an asset index proxy based on the principal components score of the indicator variables. Howe, Hargreaves and Huttly (2008) test whether different ways of weighting assets to create wealth indices are adequate proxies for consumption expenditure. In order to identify which SES proxies perform better, one must identify if empirical results differ when several different SES measurements are used.

Bollen, Glanville, & Stecklov (2002) do precisely that and use micro survey data on Ghana and Peru, acknowledging that their results may not hold for other developing countries or for alternative specifications. This paper is designed to explore their procedure using South African data. I therefore compare how the estimated impact of SES and other explanatory variables varies with the different proxies used. In addition, I show how a superior estimate of the effect of SES can be obtained by using all the proxies together.

While my analysis focuses on relatively early childbearing (birth before age 27), my methods and results can be extrapolated to fertility in general with implications extending beyond this specific dependent variable. By testing the performance of a new method of exploring socio-economic relations with fertility, I add to the evidence of the sensitivity of microanalyses to the treatment of SES.

The plan of my discussion is as follows. I first elaborate on the theoretical mechanisms and conceptualisations that underlie the relationship between SES and fertility in section 2. Section 3 presents the numerous measures of SES, while section 4 investigates how these measures are implemented in empirical studies. Section 5 reviews the literature on other comparative studies of SES proxies. I describe the data and the variables that I use in my empirical analysis in section 6. Section 7 outlines my linear probability model analysis, and presents and discusses my findings. I introduce the Lubotsky-Wittenberg regression weighted index in section 8 and explore the procedure. Section 9 concludes.

2 – SOCIO-ECONOMIC STATUS AND FERTILITY

Bollen, Glanville and Stecklov's (2001) review of the literature on SES and fertility suggests that virtually all research on fertility includes socio-economic variables, yet the mechanisms behind the relation of SES to childbearing have not been precisely identified. Consequently, there are numerous views on the theoretical avenues through which socio-economic factors are seen to influence fertility. The traditional economic perspective assumes that preferences are fixed and exogenous, and changes in fertility occur through changing opportunities and constraints (Becker, 1981).

In a broad class of theories on fertility, there is an interaction between SES, education and fertility. It is therefore no accident that the standard fertility model incorporates SES. Without trying to explain the structural pathways between the many socio-economic variables and childbearing, I aim at uncovering how the effects on fertility differ according to how these variables are measured and implemented into the equation.

While it is almost impossible to exclude the role of SES from any form of fertility analysis; it is also almost impossible to precisely define SES or even what it encompasses. In a broad sense, SES refers to a combination of both economic and sociological factors or resources that influence behaviour and decision-making (e.g. education, occupation, wealth). It is also a measure of an individual's or group's position in a hierarchical social structure. Most fertility studies endeavour to identify how a higher socioeconomic status affects childbearing.

Bollen, Glanville and Stecklov (2007) find that in these fertility studies, researchers use three broad approaches in trying to capture the effect of SES on fertility. All three focus on the reduced-form relationship.[1] The first sees SES as a unitary concept, thus having a fundamental dimension that is the primary driving force behind behavioural outcomes. Accordingly, socio-economic factors act on fertility in some combined manner rather than having individual direct effects. The second viewpoint focuses on the separate components of SES as having distinct consequences on fertility. The final treatment supposes "that these socio-economic variables are components of a general variable that has an impact, but that each of these component variables also has a specific effect that is not mediated by the general variable" (p. 17).

One of the more cited unitary conceptualisations of SES is Friedman's permanent income hypothesis of which "the permanent component [as opposed to the transitory component] is to be interpreted as reflecting the effect of those factors that the unit regards as determining its capital value or wealth..." (Friedman, 1957). These include ownership of nonhuman wealth, occupation, training, ability and personality. These factors relate to long-the run flows of income. In a similar vein, the more stable and long-term aspects of SES, are of more consequence, than the transitory components. This is especially so in the context of fertility (Easterlin, 1969), as the process of bearing and parenting children is of a long-term nature.

Permanent income can therefore be viewed as a component of SES as they share many characteristics. Although SES is an idea that is significantly broader and less developed, I use 'permanent income' and 'socio-economic status' interchangeably. In addition, both SES and permanent income are theoretical concepts and are therefore not directly measurable. This is partly the cause of the lack of consensus regarding the treatment of SES when we move from theory to empirical analysis. As will be elaborated on in the following two sections, the consequence of this lack of conceptual clarity is a plethora of measures and methods that are used to implement SES into empirical studies.

3 – MEASURES OF SOCIO-ECONOMIC STATUS

Since SES cannot be observed in any objective manner, a wide range of proxy measures or indicators are used. Some proxies are viewed as separate components of SES but often have direct influences on the outcome in question.

Occupation has been asserted to be vital in studies of class and SES in industrialized societies and in some studies, a distinction between agricultural work and non-agricultural work is prevalent (Bollen, Glanville, & Stecklov, 2001). Friedman (1957) also highlighted the importance of occupation in his permanent income hypothesis. Montgomery et al. (2000) outline certain demographic studies, many of which use occupation as a SES control. The authors also note however that in many poor countries, one's occupation is not as easily defined. Employment is often transitory in nature and many individuals have incomes from both primary and secondary employments. Coupled with the relatively larger informal sectors and higher unemployment rates in developing countries, reliable occupation details are difficult to come by. I therefore exclude any further review of its use as a proxy.

Another proxy that is closely related to SES is education. The use of readily available education details in fertility studies is however debatable since they are often presumed to have an effect on fertility, not only through SES, but through learning and knowledge (Desai & Alva, 1998). Education also improves a woman's status which can have consequences for fertility and contraception (Balk, 1994). While maternal education is almost always included, Raftery, Lewis and Aghajanian (1995) illustrate that paternal education has less subtle influences on fertility and is considered a reliable proxy for household SES.

Observed income data over a specific time period (usually a single year) can be used as a proxy but has a number of drawbacks. Firstly, the current nature of the variable does not capture the long-term nature of permanent income. Observed income can vary over time as it includes transitory components such as luck, measurement error and other unanticipated shocks to income (Deaton, 1997; Lubotsky & Wittenberg, 2006). Secondly, households in developing countries often draw their earnings from multiple sources (e.g. wage, rental, transfer), and incomes from these various sources are shown to have different effects on household behaviour (Hoddinott & Haddad, 1995). Finally, income data in these countries are often unreliable and rarely collected (Hentschel & Lanjouw, 1996).

A more common, long-run alternative to income is household expenditure or consumption. Friedman (1957) suggests that consumption behaviour is primarily driven by long-term considerations. Households tend to smooth consumption across years by borrowing and saving in order to maintain a certain standard of living. Expenditure data in developing countries is, however, difficult and costly to collect and can also be error prone (Bouis, 1994). Nonetheless, the appeal of a consumption measure has made an expenditure module a standard feature of most living standards surveys in the developing world at least since the 1990s (Deaton, 1997).

The Demographic and Health Surveys (DHS), possibly the richest sources of data on childbearing in developing countries, do not collect data on expenditures but do have information on the ownership of consumer durable goods (assets) and housing quality measures (access to water, electricity etc.). This group of measures is easier to accurately collect and is more widely available than either income or expenditure data. In fact, this group is most commonly used as proxies for SES in child health and fertility studies (Bollen, Glanville, & Stecklov, 2001). Although Falkingham and Namazie (2002) note "ownership based on information from the DHS does not reflect the *quantity* nor *quality*" of the assets or housing quality measures which may well vary across different SES groups (p. 29).

Due to the wide variety of possible proxy measures, the estimated effect of SES varies significantly from study to study. The primary obstacle is the measurement error in these proxies which results in the attenuation bias (bias towards zero) of any estimates. The more recent focus has therefore been directed towards the methods used in order to better capture the effect of this socio-economic variable thus reducing this bias.

4 – METHODS USING SOCIO-ECONOMIC STATUS PROXIES

In the context of this paper, a method refers to a specific approach, or technique, used to combine or alter certain variables.[2] With regard to SES, the predominant form of analysis is to use some method of aggregation to create a single socio-economic summary measure of other SES proxy variables. This summary measure is then entered as an independent variable in a multiple regression, along with other determinants and control variables.

Education is sometimes included in years, but normally as a series of dummy variables corresponding to the particular categories (e.g. primary, high school) of the respective country's educational system. Maternal education is sometimes used as the only control for SES. Since it has potential direct effects on fertility, it is typically used in combination with a better measure (viz. some kind of economic resource such as assets or income). This enables one to isolate the direct effect of education on fertility (Bollen, Glanville, & Stecklov, 2001).

Observed income (and expenditure alike) is usually skewed and has outliers; hence the log of the variable is used as an explanatory variable since the logged version is more normally distributed. To overcome the volatility of income, a simple average of incomes over several years is claimed to be a better proxy as it represents a more stable and permanent measure (Solon, 1992). Although Lubotsky and Wittenberg (2006) show that a weighted average is better as it accounts for the fact that income earned at different points in the life cycle may not be equal measures of permanent income.

There are many ways of incorporating measures of assets and housing quality into analyses that account for SES. One approach is to use ownership of one or more assets or housing quality characteristics (such as a car, television, flushing toilet) as single indicator variables (Devin & Erickson, 1996). Another strategy is to construct an index of these variables. This can be an indicator variable for whether a household owns any one of a set of items (Muhuri, 1996), although creating an equally weighted index is far more common (Bollen et al., 1995; Gorbach et al., 1998).

The simplest form of this type of index is to sum the number of types of assets owned. While easy to compute from the simple asset indicators available in most datasets, this method offers little theoretical value as it places relatively inexpensive items, such as radios, on the same scale as more valuable assets like cars. "It can also have paradoxical effects when certain assets are 'inferior goods', so that their ownership makes households look more affluent when in reality it might signal less affluence" (Wittenberg, forthcoming). Further complications arise when trying to include housing quality measures (such as type of toilet or water supply) where more than two options usually exist.

The use of a principal component analysis (PCA) to create the weights for the asset index has been given much attention of late (Filmer & Pritchett, 2001; McKenzie, 2005; Vaya & Kumaranayake, 2006). These weights can be thought to represent the internal structure of the set of variables in a way that explains the greatest proportion of the total variation between all the variables in the set. Assuming that SES explains the maximum variation in the asset variables, PCA is a pragmatic choice to a data constraint problem (Filmer & Pritchett, 2001). Although a principal component based index may incur "clumping and truncation issues", these issues do not affect SES measures as much as measures of inequality (McKenzie, 2005, p. 230).

Less common weighting methods have also been employed. Dargent-Molina et al. (1994) use the monetary value of each asset; however reliable estimates of the value of such goods are not widely available. Difficulties also arise in taking into account and adjusting for regional price variations and inflation. Some claim a multiple correspondence analysis (MCA) obtains more appropriate weights for categorical variables than PCA, but do not find much difference in their results (Booysen et al., 2008; Howe, Hargreaves, & Huttly, 2008).

The use of a wide variety of variables, in an array of different methods, by multiple researchers, all attempting to control for SES – suggests that these variables "tap the same underlying concept" (Bollen, Glanville, & Stecklov, 2001, p. 163). Following up this idea, Bollen, Glanville and Stecklov (2007) strive to formulate a more realistic model that relates the components of SES, including permanent income, to fertility in developing countries. They treat permanent income as the latent variable and use structural equation modelling to control for the confounding effects of measurement error that is intrinsic to the use of proxies.

Structural equation models (SEM) allow some variables to be latent; therefore the coefficients can be interpreted as effects to and from the various combinations of latent and observed variables. These coefficients are essentially regression coefficients. Although Bollen, Glanville and Stecklov (2007) argue that "SEM is less restrictive and more complex than multiple regression in that there are multiple indicators of the latent variables, multiple equations, and [one can] include random and non-random measurement error" (p. 23).

The authors' results suggest that the estimated negative effect of permanent income on childbearing in Ghana and Peru is much higher when treated as a latent variable, as opposed to using imprecisely measured proxies. In addition, they find that the effect of other components of SES, such as occupation and education, is largely mediated through permanent income. In this context, a proxy variable is used to represent the unobservable variable of interest, SES. Structural equation modelling treats SES as a latent variable removing the use of any SES proxies in the estimated equations. SEM therefore lies beyond the scope of this paper.

Finally, there is one other method capturing SES that has not been implemented in fertility studies before. This method relies on the Lubotsky-Wittenberg estimation procedure (Lubotsky & Wittenberg, 2006). It involves entering all the proxies individually into the regression and aggregating afterwards. Their procedure is "best thought of as a method to interpret the coefficients in a regression under the null hypothesis that the variables are all generated by a common latent factor" (p. 550). Given the discussion of SES (and permanent income) above, as well as the results of Bollen, Glanville and Stecklov's (2007) latent variable approach; it is not unreasonable to assume that this null hypothesis does in fact hold. I elaborate on this method in section 8.

5 – COMPARATIVE STUDIES OF SOCIO-ECONOMIC STATUS PROXIES

Although there is no agreement on the optimum measures or methods to use, there are other studies that compare the performance of different proxies under several behavioural and health outcomes of interest.

Montgomery et al. (2000) use data on five developing countries and compare three ways of using asset indicators (simple-sum, simple-sum values as separate indicator variables, and individual dummy variables for each asset) on fertility, child mortality and children's schooling. They conclude that using assets to create a standard of living index leads to weak predictions of consumption. Nevertheless, they are still worth considering for their use in testing the hypothesis that the estimated effect of permanent income is zero. They also note that certain measures, like having access to piped water also have direct effects on mortality. Certainly this is intuitive. These proxies are consequently performing two tasks with effects that are difficult to separate.

Filmer and Pritchett (2001) evaluated and promoted the use of principal component analysis (PCA) in creating an index, although "this approach has been used for over a decade to proxy the level of socioeconomic status in the child development and nutrition literature" (McKenzie, 2005, p. 232). They found that as a proxy for wealth, this PCA index was more consistent over time and a better predictor of school enrolment in India when compared to that of household expenditure.

Furthering the study of the use of a principal components score index, Houweling, Kunst and Mackenbach (2003) test the sensitivity of this index to the type of assets included. They provide empirical evidence on under-5 mortality and measles immunisation coverage for ten developing countries, using DHS survey dataset. They conclude that "users of an asset index should, however, be aware that choice of assets influences the outcomes observed". In accordance with Montgomery et al. (2000), they show that for explanatory purposes, certain sets of assets and housing quality measures should be used with caution as one may not be able to disentangle their direct effects on the outcome of interest from those mediated through SES.

Howe, Hargreaves and Huttly (2008) recently shed some light on how the coding of the assets and housing quality measures (viz. dichotomised or categorical) in datasets affects the performance of these indices. Using data from Malawi, they compare the difference between five asset based wealth indices that are used as proxies for socio-economic position (SEP) in epidemiological studies. While all demonstrated "similarly modest agreement" with consumption expenditure, they found that "the indices constructed using dichotomised data showed strong agreement with each other, as did the indices constructed using categorical data" (p. 1).

A comprehensive comparison of several different proxies for economic status (including occupation, expenditure and variations of four asset indices) was undertaken by Bollen, Glanville and Stecklov (2002). In the context of fertility in Ghana and Peru, the authors found a principal component analysis to have the strongest effect, followed by a simple sum measure. They also find that much of the effect of maternal education "is through its direct effect on fertility rather than as a proxy for economic status" (p. 88).

Although my literature review is far-reaching, it reveals no consensus on any of the above issues. The underlying mechanism through which SES influences fertility differs from one behavioural perspective to another which naturally affects how SES is then measured and used. Despite the data constraints, there is no measure of permanent income that is seen to stand out above the rest, besides perhaps consumption expenditure. Each method used to assimilate the many asset variables available has significant shortcomings subject to the outcome of interest, choice of assets and coding of data. Most studies report inconclusive evidence as to which measure or method extracts and captures the most information.

I therefore extend such research by using similar comparison procedures in the context of fertility. I first examine the performance of several of the SES proxies that have been reviewed and how they affect other explanatory variables. I then analyse the use of a new estimation method and test its versatility in providing better estimates of the elusive effect of SES.

6 – DATA

The latest wave (Wave 4) of the Cape Area Panel Study (CAPS) dataset will be used in my empirical work.[3] CAPS is a longitudinal study of the lives of youths and young adults in metropolitan Cape Town, South Africa. Using a stratified two-stage random design, the first wave sampled 4800 randomly selected young people between the ages of 14 and 22 in August-December (2002) through interviews. The youth sample had been interviewed again in 2003-04 (Wave 2), 2005 (Wave 3) and 2006 (Wave 4) (Lam, et al., August 2008). The study covers: a wide range of outcomes including schooling, employment, health, family formation, and intergenerational support systems; as well as household characteristics such as income, expenditure, consumer durable goods and housing quality measures. Essential fertility information and data for constructing a diverse set of SES Proxies is therefore available.

Many fertility studies examine the household characteristics and behaviour of women of all ages who are cohabiting with their spouses, thus forming the typical nuclear family. Consequently, any socio-economic measurements (regardless of the method employed) largely reflect a women's *current* SES, which may well have changed from when she had had her first child and would therefore not be entirely relevant. As Bollen, Glanville and Stecklov (2007) note "examining births within a relatively short time period allows us to better evaluate the influence of SES and permanent income" (p. 19). The CAPS dataset allows one to lock onto this issue simply because of the young age group of the sample.

Many of the young women do not assume the role of the female household head in a typical nuclear family. In fact, over 85 percent of the young women are neither the heads of the households nor wives, partners or girlfriends to the heads of the households. Many of them share a home with their parents, grandparents, cousins and other relatives. Over 45 percent form part of a household where the household size ranges from 6 to 20. The CAPS dataset therefore encompasses socio-economic information on the households in which many of the young women have been raised. While SES is hypothesised to be a determinant of fertility in general, in the case of *young* women it is the socio-economic *background* that should play a role in childbearing. The unique setup of the dataset therefore provides the opportunity to analyse the background of many young adults closer to when they may have first given birth, as opposed to simply using their current status.

I do not use the data as a panel but rather as a cross-section hence the details regarding attrition are of little relevance. The 2006 wave, wave 4, included 71.9 percent of the original sample in 2002 i.e. 3451 young adults. South Africa is a racially diverse country with a minority of White and Indian households. As a result 267 young adults were dropped as they belonged to White or Indian households in which the percentage of young adult mothers was close to zero. A further 82 observations had to be removed as they had missing fertility information and a further 1400 males were removed. Finally 72 observations were dropped as those young adults fell outside the age range under investigation of 18-26 year olds. This yields a total of 1630 females between the ages of 18 and 26 living in African or Coloured households in the Cape Town Metropolitan area.

6.1 – Definitions of Variables

Dependent Variable – I use the indicator variable of having had a 'live birth ever' as the dependent variable given that there were almost no still births or miscarriages in the sample. While Bollen, Glanville, & Stecklov (2002) use the number of 'children ever born'; the choice of birth ever is more appropriate in this sample as only 9.8 percent of the mothers have had two children with 1.4 percent having had three. The young age of the sample does not allow the explanatory variables that affect a woman's fertility to vary much over time. Therefore the birth ever variable reflects a relatively current association between these explanatory variables and fertility.

Control Variables – Bollen, Glanville and Stecklov (2002) note that a "basic model of fertility" typically includes a set of exogenous control variables (p. 85). In my reduced-form models for the young adult sample of Cape Town I use age, population group and marital status. The inclusion of age is based on the hypothesis that the older a woman is, the more likely she is to have had a child at any point in time. Whereas the population group control is used because Africans and Coloureds may differ with respect to bearing children (Moultrie & McGrath, 2007), as they do for many of the explanatory variables outlined below. Finally having a child is more likely if one had been married at some point. Cross tabulations of birth ever with each control variable confirmed sufficient observations within each cell. These variables, along with education variables, are described in Table 1.

Table 1 – ***Description of Control Variables in the 2006 Cape Area Panel Study (CAPS)***

Young Adults Age	18 (reference), 19, 20, 21, 22, 23, 24, 25 and 26
Population Group	African (reference) and Coloured
Never Married	Equals 1 if the young adult has never married Equals 0 if married, divorced, separated or widowed
Young Adult's Completed Education	Up to grade 9 (reference), grade 10 or grade 11, grade 12 and higher than grade 12

The variables in Table 1 form the basis of my model and I introduce a series of SES measures, each of which are described below. Since the income and expenditure proxies are skewed and have outliers they are all logged (first adding a constant to the measures to make the lowest value equal to one).

Income – The log of monthly per capita income from wave 1 and wave 3 are used as two proxies that are viewed as transitory. Utilising the time or panel dimension of the CAPS dataset, I include the average income over the two years as a more long-term measure (i.e. average of the two logs of monthly per capita income). It must be noted that wave 3 income data did contain some missing values. These missing data were imputed. The details of all imputations are described in appendix A1.

Expenditure – The expenditure module in the survey is a short but comprehensive model that is near identical to many other commonly used surveys in South Africa. The log of household expenditure is a typical proxy for SES. I divide total monthly expenditure by the number of *people* in the household before logging it. It is common to divide by the number of *adults* in the household so as not to create a feedback between the denominator and a component of the dependent variable (usually number of children). However, the unconventional use of this unique young adult sample (as discussed above), coupled with the dichotomous birth ever variable, dampens the confounding effect of expenditure per person.

Household Consumer Durable Goods (Assets) – CAPS provides details on the ownership of 14 assets: radio, television, video/DVD player, landline, cellular phone, refrigerator/freezer, gas or electric stove, microwave, washing machine, bicycle, motorcycle, motor vehicle, computer/laptop and more than 5 books.

Two indices that are commonly used in the literature are also created here using these assets: a simple-sum (SS) and a principal components analysis (PCA) score. The former is created by summing the number of the types of assets that a household owns therefore weighting each type of good equally. The principal components method assigns each type of good a weighting based on the principal components score, as used by Filmer & Pritchett (2001). The first component of the full set of assets (eigenvalue of 4.6279) captured about 33 percent of the variation in the data. The particulars are outlined in appendix A2.

Another set of indices are created corresponding to the same two procedures but using a more restrictive set of assets (viz. radio, television, refrigerator, bicycle, motorcycle and car) that are found in most of the Demographic and Health Surveys (DHS).

Housing Quality (HQ) – Finally, the four indices used above are modified to include housing quality measures that are commonly used in models accounting for SES. These measures include the indicator variables of a brick-type house, presence of a flushing toilet, piped-water and electricity, as well the number of rooms per person.

Table 2 reports descriptive statistics for all variables used in the regressions. Appendix A3 contains descriptive statistics for all original variables that were used to create the proxies.

Table 2 - *Regression Variable Descriptive Statistics in the 2006 Cape Area Panel Study (CAPS)*

Regression Variables	Mean	Std. Dev.
Dependent Variable		
Birth Ever	0.4312883	0.4954082
Socio-economic Status Proxies		
Log Per Capita Income wave 1	5.954975	1.0297340
Log Per Capita Income wave 3	6.344040	1.0355910
Average Log Per Capita Income	6.109465	0.8612216
Log Expenditure Per Person	5.913113	0.8808194
Simple-Sum (SS)	7.650307	3.1863230
Principal Components (PCA)	0	2.1512550
Simple-Sum Including Housing Quality (SS+HQ)	12.019880	4.0393560
Principal Components Including Housing Quality (PCA+HQ)	0	2.4196500
DHS-Simple-Sum (DHS-SS)	3.039264	1.1646230
DHS-Principal Components (DHS-PCA)	0	1.4350570
DHS-Simple-Sum Including Housing Quality (DHS-SS+HQ)	7.408837	2.0926000
DHS-Principal Components Including Housing Quality (DHS-PCA+HQ)	0	1.9056390
Education		
Up to grade 9 (reference)	0.2742331	0.4462639
Grade 10/11	0.3441718	0.4752432
Grade 12	0.3417178	0.4744311
Greater than Grade 12	0.0398773	0.1957310
Age		
18 (reference)	0.1030675	0.3041403
19	0.1251534	0.3309943
20	0.1300613	0.3364741
21	0.1306748	0.3371478
22	0.1300613	0.3364741
23	0.1134969	0.3172966
24	0.1061350	0.3081048
25	0.0901840	0.2865332
26	0.0711656	0.2571802
Population group		
African (reference)	0.5171779	0.4998582
Coloured	0.4828221	0.4998582
Marital Status		
Never Married	0.8852761	0.3187863
N=1630		

7 – THE LINEAR PROBABILITY MODEL

I use a linear probability model (LPM) which is simply an ordinary least squares (OLS) regression run on a binary dependent variable, in my case, birth ever[4]. I implement standard Huber/White techniques to correct standard error estimates resulting from the cluster design of the sample. Using population group as a control variable allows me to account for the stratification of the survey design as the CAPS dataset is stratified by population group. I account for sample design, and both household and young adult non-response by using the weights provided.

Closely following Bollen, Glanville, & Stecklov (2002), my initial model includes only the controls, namely age, race and never married. Education is often found to have direct effects on fertility and the education variables are added accordingly thus creating the second model. Since education is closely related to SES, this model can be seen as baseline against which I can compare changes in the parameters for these variables as a result of adding the different SES proxies. I therefore proceed to add each SES proxy independently in order to assess the performance of each one. As mentioned above, the focus is not on the actual effects on fertility but rather how the effects differ according to which proxy is used.

I evaluate the performance of the different SES proxies in several ways. In gauging the effect of the different measures of SES on the control variables, I identify any shifts from significance to insignificance (or vice versa). I also note the changes between the maximum and minimum values of the parameters for those controls that are statistically significant. For the proxies, I examine the statistical significance; compare (where possible) the magnitude of the un-standardised and standardised coefficients; and evaluate the relative fits by comparing the R-Squared values. The same sample is used in each model in order to rule out sample differences. I set up a final comparison in the section 8 using the Lubotsky-Wittenberg method.

7.1 – Comparing the Effects on the Control Variables

Table 3 reports the regression results of *most* of the models for the young adult sample of Cape Town. To conserve space, I report the results of *all* the models in a later table that compares all the proxies. All my subsequent results for the control variables do however, apply to all the models and I extend my analysis to those cases where they do not.

All the exogenous controls of model 1 are significant except for Coloured. Age has a strongly positive and significant effect on fertility hence the probability of having a birth generally increases with an increase in age. Having never been married significantly reduces the chances of having a birth.

Model 2 includes the education dummies of which all are significant. The probability associated with having a birth decreases as the education level of the woman increases. Adding education variables cause the coefficients on the age variables to increase. This increase is especially evident in the higher age groups where the expected probabilities increase by more than 5 percent. Since it is more likely that education is correlated with age in this particular sample, this result suggests that controlling for education reveals a 'truer' age effect on fertility.

Models 3 to 8 illustrate the effects of using different proxy measures for SES. Each proxy variable has a significant effect on fertility, net of the education and control variables, and serves to reduce the probability of having a birth as the variable's value increases. My results differ slightly from Bollen, Glanville, & Stecklov (2002) in that the log of monthly expenditure per person has the largest effect, but I reserve a comparison of the proxies for the next section.

Including a measure of SES in the models, alters the coefficients of some of the explanatory variables. Coloured becomes significant only in model 4 when expenditure is the chosen proxy. When compared to model 2, all the education coefficients in each proxy model decrease (especially those of the higher education levels); with the highest decrease in all categories occurring when expenditure is the proxy under consideration. Therefore, in contrast to the findings of Bollen, Glanville, & Stecklov (2002), mine illustrate that women's education has some mediating effect on fertility through SES, as opposed to having only a direct effect. Of course this difference can be attributable to the different samples under investigation.

Table 3 - *Proxy Results based on the 2006 Cape Area Panel Study (CAPS), Sample size: N=1630*

	Model 1		*Model 2*		*Model 3*		*Model 4*		*Model 5*		*Model 6*		*Model 7*		*Model 8*	
	Controls Only		Plus Education		Average Income		Expenditure		SS		PCA		SS + HQ		PCA + HQ	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
SES Proxies																
Average Income					**-0.071620**	**(0.0163)****										
Expenditure							**-0.127253**	**(0.015)****								
Simple-Sum									**-0.020806**	**(0.005)****						
PCA											**-0.031134**	**(0.007)****				
Simple-Sum + HQ													**-0.020045**	**(0.004)****		
PCA + HQ															**-0.028034**	**(0.006)****
Education																
Grade 10 or 11			-0.136189	(0.031)**	-0.109842	(0.031)**	-0.081815	(0.031)**	-0.107318	(0.032)**	-0.108879	(0.032)**	-0.100434	(0.032)**	-0.109572	(0.032)**
Grade 12			-0.291254	(0.030)**	-0.241178	(0.032)**	-0.189709	(0.032)**	-0.238334	(0.033)**	-0.240523	(0.033)**	-0.227721	(0.033)**	-0.244082	(0.032)**
Greater than 12			-0.405736	(0.050)**	-0.329325	(0.052)**	-0.250464	(0.051)**	-0.327939	(0.053)**	-0.329406	(0.053)**	-0.314254	(0.053)**	-0.336443	(0.052)**
Age																
19	0.112276	(0.044)*	0.140556	(0.043)**	0.131979	(0.044)**	0.125766	(0.043)**	0.131169	(0.043)**	0.131485	(0.044)**	0.130781	(0.040)**	0.133547	(0.044)**
20	0.139739	(0.043)**	0.191022	(0.043)**	0.181875	(0.042)**	0.168066	(0.042)**	0.181522	(0.042)**	0.181632	(0.042)**	0.181506	(0.042)**	0.183707	(0.043)**
21	0.220406	(0.045)**	0.267887	(0.045)**	0.254761	(0.046)**	0.238852	(0.045)**	0.255301	(0.045)**	0.255293	(0.045)**	0.255462	(0.045)**	0.258990	(0.045)**
22	0.167045	(0.044)**	0.258831	(0.045)**	0.247171	(0.044)**	0.232243	(0.044)**	0.239248	(0.045)**	0.239948	(0.045)**	0.238305	(0.045)**	0.242643	(0.045)**
23	0.263610	(0.047)**	0.332143	(0.046)**	0.334518	(0.047)**	0.329625	(0.046)**	0.322025	(0.046)**	0.322641	(0.046)**	0.322029	(0.046)**	0.323244	(0.046)**
24	0.423852	(0.049)**	0.494560	(0.049)**	0.489315	(0.048)**	0.471595	(0.048)**	0.480670	(0.048)**	0.481245	(0.048)**	0.478476	(0.048)**	0.482000	(0.049)**
25	0.479225	(0.051)**	0.542970	(0.049)**	0.533318	(0.049)**	0.524374	(0.047)**	0.525959	(0.049)**	0.526339	(0.049)**	0.525021	(0.049)**	0.529683	(0.049)**
26	0.455752	(0.056)**	0.525636	(0.053)**	0.522556	(0.053)**	0.522687	(0.053)**	0.505389	(0.053)**	0.506558	(0.053)**	0.502879	(0.053)**	0.508244	(0.053)**
Population Group																
Coloured	-0.040611	(0.023)	-0.025232	(0.023)	0.040383	(0.027)	0.072815	(0.025)**	0.057577	(0.030)	0.060156	(0.030)*	0.075950	(0.03)*	0.058864	(0.030)*
Marital Status																
Never Married	-0.388486	(0.035)**	-0.356110	(0.035)**	-0.360960	(0.035)**	-0.372388	(0.033)**	-0.357893	(0.035)**	-0.357490	(0.035)**	-0.353762	(0.035)**	-0.353594	(0.035)**
Constant	0.555180	(0.049)**	0.630979	(0.053)**	1.020156	(0.099)**	1.310316	(0.093)**	0.734217	(0.056)**	0.574377	(0.055)**	0.797970	(0.060)**	0.571291	(0.060)**

Note: Robust standard errors in parentheses. * significant at 5%; ** significant at 1%

The dependent variable is an indicator that a woman has had a birth.

The next logical step would be to compare whether the coefficients change when the different proxies are used as opposed to adding a particular proxy. Table 4 presents the percentage difference between the largest and smallest estimated effects for each control variable that is statistically significant for models 3 to 8. It is evident that the parameters for the education variables are affected much more, by the choice of proxy, than those of the age variables. It is interesting to note that expenditure features in almost all coefficient differences as either the maximum or minimum, whereas the principal components (including housing quality) and the average income proxies usually have the largest coefficients.

Now I go on to extend the comparison to include the results for those models not reported in table 3 (viz. the individual wave incomes and the DHS subset proxies). The results (not reported) indicate that the DHS subset principal components (including housing quality) model produces even larger coefficients on the control variables than any of those indicated in table 3 thus leading to even greater percentage differences. Overall, any general conclusions of the impact of maternal education on fertility would therefore be influenced by the choice of proxy used.

Table 4 - ***Percentage Difference Between the Maximum and Minimum Coefficients based on the 2006 Cape Area Panel Study (CAPS)***

Control Variable	Model with Maximum Coef.	Model with Minimum Coef.	Percentage Difference (max-min)/min
Education			
Grade 10 or 11	Average Income	Expenditure	34.3%
Grade 12	PCA + HQ	Expenditure	28.7%
Greater than 12	PCA + HQ	Expenditure	34.3%
Age			
19	PCA + HQ	Expenditure	6.2%
20	PCA + HQ	Expenditure	9.3%
21	PCA + HQ	Expenditure	8.4%
22	Average Income	Expenditure	6.4%
23	Average Income	SS	3.9%
24	Average Income	Expenditure	3.8%
25	Average Income	Expenditure	1.7%
26	Expenditure	SS + HQ	3.9%
Population Group			
Coloured	*Not significant		
Marital Status			
Never Married	Expenditure	SS + HQ	2.8%

7.2 – Comparing the Effects of the SES Proxies

Table 5 displays the un-standardized and standardized coefficients for all the proxy variables as well as the R-Squared values and rankings for each respective model.

Table 5 - ***Comparing Proxies for SES based on the 2006 Cape Area Panel Study (CAPS)***

Proxy	Un-standardized Coefficient	S.E.	Standardized Coefficient	R^2	R^2 Ranking
Income (wave 1)	-0.0573858	(0.012575)**	-0.1198209	0.2757	4
Income (wave 3)	-0.0392777	(0.014822)**	-0.0771578	0.2705	9
Average Income	-0.0716204	(0.016260)**	-0.1273063	0.2761	3
Expenditure	-0.1272534	(0.015046)**	-0.2267243	0.2997	1
SS	-0.0208064	(0.005022)**	-0.1321128	0.2747	6
PCA	-0.0311342	(0.007472)**	-0.1322118	0.2745	7
SS + HQ	-0.0200448	(0.003898)**	-0.1589977	0.2784	2
PCA + HQ	-0.0280336	(0.006395)**	-0.1300519	0.2745	7
DHS-SS	-0.0269568	(0.011494)*	-0.0621837	0.2687	11
DHS-PCA	-0.0203106	(0.009210)*	-0.0555257	0.2682	12
DHS-SS + HQ	-0.0300493	(0.006513)**	-0.1202384	0.2750	5
DHS-PCA + HQ	-0.0189642	(0.006801)**	-0.0654691	0.2690	10

Note: Robust standard errors in parentheses. * significant at 5%; ** significant at 1%

The first thing to note is that all coefficients are significant and negative at the 1 percent level except for two of the DHS-set proxies (the two without the housing quality measures) which are only significant at the 5 percent level. The three income proxies as well as the expenditure measure are logged thus generating semi-log models. Each un-standardised coefficient multiplied by 0.01 can therefore be interpreted as the expected difference in probability of a birth for a one percent change in the unlogged proxy variable. In comparing these four measures, expenditure has the highest estimated impact, followed by average income.

Interpretation of the remaining un-standardized coefficients is not as straight forward since I am comparing across equations with variables that have different units of measurements. To overcome this difficulty, I compare the standardized coefficients (Gujarati, 2003) listed in the fourth column[5]. These can be understood as the difference in number of standard deviations to expect in the birth ever variable for a single standard deviation shift in the explanatory variable.

Four striking observations can be noted from this column. Firstly, the log of expenditure per person clearly has the 'strongest' effect. Secondly, while the observed income from wave 1 has a strong effect, the average income proxy has higher coefficients than both of the individual wave income measures. It does however fall short when compared to any of the full-set asset proxies. The third observation is that each full-set asset index coefficient is higher than its corresponding DHS set estimate. Fourthly, both the full-set and DHS-set simple-sum measures that include housing quality are noticeably higher than the other asset proxies in their respective sets.

I finally evaluate the relative fit of each of the models by comparing the standard R-squared values. Since each model has the same sample size and the same number of regressors, there is no benefit in using the adjusted R-squared statistic as the ordering of the values will remain the same. Once again the model with expenditure is ranked first, explaining the greatest proportion of the total variance of the data than any other model. The next best predictor is the full-set simple-sum proxy that includes housing quality, closely followed by the average income and wave 1 observed income proxies.

7.3 – Discussion of the Proxy Results

Consistent with theory and other empirical studies, a higher SES leads to a lower probability of having a child. While it is clear from all the evaluation techniques shown in table 5 that expenditure is the best empirical reflection of SES when compared to any of the other proxies, it is sometimes found to be endogenous in fertility models thus having a positive impact (Bollen, Glanville, & Stecklov, 2002).

So why does this endogeneity of expenditure appear to be muted, or even 'reversed' in this analysis? A possible explanation is based on the hypothesis that in this young sample, it is the socio-economic *background* (more so than the usually defined SES) of a young woman that is a determinant of her childbearing behaviour. As mentioned in section 5, many of these young women still live with their parents or under other family member's guardianship (85.5 percent). The proxies are therefore picking up the signal stemming from the socio-economic background of the majority of the young women.

Assuming this hypothesis is true, then the results indicate that household expenditure captures this 'background effect' the best. What makes expenditure so different? Since most of the young women are not the primary income earners or consumption decision-makers of the household, they do not have command over the household resources. This means that the household expenditure measure better reflects the SES of the *family* and household of which a particular young adult is simply part of. Hence the expenditure variable better reflects her socio-economic background, as opposed to simply being correlated with the number of children a mother may have.

One could argue that this so called 'background effect' should then apply to all the proxies. Indeed this may be true, especially for the better proxies (see appendix A5). However, consumption expenditure has always been the benchmark to which other proxies have been compared and for good reasons. While in many other studies mentioned above, some of these other proxies can be placed in the same league as expenditure, but for reasons argued below, my results prove that they simply cannot compete.

Observed income from wave 1 performed surprisingly well considering the above discussion on the likelihood of measurement error. Since wave 1 data was obtained four years prior to wave 4, a larger proportion (94.5 percent) of the young women were still living with their parents or other family members. The regression coefficient is therefore capturing the 'background effect' of higher number of young adults and if this effect on childbearing is substantial, the estimate would be inflated. The relatively good performance of wave 1 observed income is therefore not necessarily indicative of the income measure being a good proxy. The poor performance of the wave 3 income proxy confirms this. Both measurement error and imputation related issues are likely causes for its poor performance.

Despite the fact that the average income proxy is partly based on wave 3 income, it has a higher R-squared value and higher coefficients (both un-standardised and standardised) than either of the individual wave income proxy models suggesting that the problem of attenuation is reduced. Although confirmation from a greater than two-year average would be beneficial, this result is consistent with studies indicating that an average of income over several years is more of a precise measure of permanent income than a single year's observed income (Lubotsky & Wittenberg, 2006; Solon, 1992).

Regarding the asset indices, the four DHS-set indices clearly do not perform as well as their full-set counterparts. This confirms that a larger and wider set of assets do improve the performance of asset indices, especially those based on PCA (McKenzie, 2005). Regardless of this result, none of the asset-based approaches stood out significantly as some of them did in Bollen, Glanville, & Stecklov's (2002) comparison. This can be primarily attributable to the sample under investigation whereby not much variation in asset ownership occurs across households.

If one looks at the mean and standard deviation of each indicator variable in appendix A3, it becomes clear that a high percentage of the households (greater than 80 percent) either have, or do not have, the particular asset or housing quality measure under consideration. Certainly for the PCA, it is those assets exhibiting more of an unequal distribution across households that are given more weight. Therefore variables with a low standard deviation will be given a low weight as those particular assets do not facilitate the differentiation of SES levels (McKenzie, 2005).[6]

While adding continuous variables to a PCA may improve the behaviour of the index (Vaya & Kumaranayake, 2006), this would certainly warp any simple-sum calculation. The only continuous variable (number of rooms per person) features in the housing quality measures. Consequently, both the mean and standard deviation of both the full-set and DHS-set simple-sum proxies that do include the housing quality measures are higher than their respective counterparts that do not (see Table 2). This would influence the calculation of the standardised coefficients and the results in Table 4 do appear to reflect a possible distortion. Howe, Hargreaves and Huttly (2008) attest that the choice of data included in calculating an index has a larger impact than the actual weighting method used. It is likely that this is the case here too. The end result is that consumption expenditure best utilises the unique construct of this dataset to proxy for SES.

7.4 – Implications of the Results

Thus far we have established that SES plays a complicated role in fertility analysis. There is a large degree of uncertainty over how to define, measure and incorporate SES into empirical work. What is certain is the fact that a higher level SES has a negative effect on childbearing. This is evident in that in all the models, the SES proxy coefficients are significant (and negative). Despite this significance, there exists a huge discrepancy between the magnitudes of the coefficients of both the covariates and the SES proxies across models.

With respect to the covariates, if the choice of proxy alters the coefficients of certain covariates, then a researcher's use of these coefficients in putting forth any arguments in regards to the role of these explanatory variables should be scrutinized. More specifically, if maternal education is the variable of interest, then the choice of SES as well as the sample under consideration will influence the results and hence any economic conclusions.

Inherent in use of proxy variables is the difficulty of separating the direct effects on fertility of the variable from those mediated through SES. In terms of the role of SES on childbearing behaviour, researchers may find statistically significant and robust results, but whether these results can actually be used to infer the effect of SES on the observed differences in fertility is debateable. Any policy recommendations based on studies aimed at identifying the determinants of fertility should therefore be approached with caution.

8 – THE REGRESSION WEIGHTED INDEX

It clear that using proxies for an abstract concept like SES will always lead to an attenuation bias due to the inherent measurement error. Regardless of how well a particular proxy like consumption expenditure may perform under certain circumstances, there is no *single* optimum proxy that can be used across-the-board to reliably account for SES. This does not imply that we cannot achieve the best possible estimate for the effect of SES by using all the information available at our disposal. Lubotsky and Wittenberg (2006) provide an estimation method that can be used "for the case where a researcher has multiple proxies for a single, unobserved independent variable" (p. 559).

They prove that under certain conditions the attenuation bias is minimized by entering all the proxies in a regression simultaneously and then constructing a weighted sum of the coefficients on each proxy variable. Hence the Lubotsky-Wittenberg (LW) estimator can be written as

$$\hat{\beta}_{LW} = \hat{\rho}'\hat{b} = \sum_{i=1}^{k} \hat{\rho}_i \, \hat{b}_i \qquad (1)$$

where b_i is the coefficient on the x_i proxy in the multiple regression including all the proxies. The ρ vector can be consistently estimated by the generalised method of moments estimator

$$\hat{\rho}_i = \frac{cov(y,x_i)}{cov(y,x_1)} \tag{2}$$

where y is the dependent variable.

When applied to fertility analysis, the LW estimation procedure gives a way to interpret how changes in the underlying unobservable variable, SES (or permanent income), affects childbearing. Another advantage of this method is that it implicitly constructs a composite linear index based on the individual proxies. The individual proxy coefficients are the weights and the $\hat{\beta}_{LW}$ is the coefficient on the index, hence the name 'regression weighted index'. Lubotsky and Wittenberg (2006, p. 554) show that the index can be written explicitly as

$$a_{LW} = \frac{1}{\hat{\beta}_{LW}} \sum_{i=1}^{k} a_i \, \hat{b}_i \tag{3}$$

It must be noted that "this index is not an all-purpose income or welfare proxy" and is subject to the dependent variable under consideration (Wittenberg, forthcoming). In the context of this study, this index would capture that part of the common correlation between the SES proxies that best explains the probability of having a child. An attractive feature of this index is that its coefficient $\hat{\beta}_{LW}$ can be compared to the coefficient on any other linear combination of proxies. These include those linear combinations created by: a principal component analysis; a simple-sum of the proxies; or an individual proxy variable used on its own (for example, how observed income or expenditure is often used). Lubotsky and Wittenberg (2006) note that, before any comparisons can be made, the coefficients on the other linear combinations need to first be placed on the same scale.

8.1 – Employing the LW Estimation Procedure

The first step is to fix the scale of the latent variable in terms of the observable x_1 such that $\rho_1 = 1$. This scaling is essential when comparing estimates across models. In my analysis I have chosen to standardise my estimates on the wave 1 observed income proxy. There are two reasons for this choice. Firstly the bivariate correlation between this variable and the dependent variable, birth ever, is relatively strong. More importantly, a higher observed income level is usually associated with a higher SES, which in turn impacts childbearing. Although observed income is a noisy measure of permanent income, there are no other pathways for the variable to affect fertility other than functioning as a SES proxy. Indeed, the validity of the LW procedure relies on the assumption that all the proxies used in the regression are in fact proxies for the same latent variable and do not actually have any direct effects.

Although adding more proxies will strengthen the estimate of the regression weighted index, if some of the proxies do belong in the main regression then this would "contaminate the interpretation of the index" (Wittenberg, 2007, p. 4). The proxies in my analysis are the two observed income wave variables, expenditure and all the individual assets and housing quality measures. Given the results and discussion in the previous section; it is reasonable to assume that the expenditure proxy behaves in a similar manner to observed income in that neither imposes any effects other than those mediated through SES. The asset variables however, require further investigation.

Wittenberg (2007) offers a set of procedures that can test for whether a set of variables do all proxy for a common latent variable. This specification test has potential to further the understanding of the underlying relationships between fertility and SES, as well as those between the SES proxies. This paper is largely comparative, and not explanatory, in nature so, with the aid of two rudimentary assessments, I instead follow the author's remark "these techniques are no substitute for thinking about the underlying relationships theoretically" (p. 27).

The most basic 'test' for any relationship is to simply look at the raw correlations and gauge whether there is a strong link or not. One can see from table 6, aside from the number of rooms per person and owning a computer, none of the assets have particularly strong correlations with the BE variable besides the A computer is a relatively expensive item, especially in developing countries and it is much more likely that owning one would indicate a higher SES as opposed to directly affecting the probability of having a child. One could argue that the higher the number of rooms per person leads to more privacy resulting in more sexual activity which could increase the probability of having a child. However the negative correlation coefficient on this proxy suggests that a higher number of rooms per person indicate a larger house which can then be related to a higher SES.

Another assessment is to run a regression with all the asset variables in addition to the income and expenditure proxies. Since the income and expenditure proxies can be seen as 'pure' SES controls, the statistical significance of any of the asset coefficients could indicate that they belong in the main regression and should not be aggregated in the index. The results from proxy set 1 in table 6 indicate that owning a TV, bicycle, electricity as well as the number of rooms per person could have direct effects on the probability of having a child.

However, the use of income and expenditure in the regression as SES proxies leads to estimation in the presence of measurement error. Consequently the coefficients on these assets may be statistically significant only because the assets are picking up part of the signal of the latent variable (SES) as opposed to indicating a direct effect (Wittenberg, 2005). This latter argument seems more plausible. The percentage of households that own a TV or have electricity (as indicated in appendix A3) is fairly large while that that owns a bicycle is rather low. This raises doubt as to whether these variables vary enough over households to actually produce direct effects especially since each variable's correlation with the dependent variable is rather low. The conclusion is that it is unlikely for any of these asset variables to serve any other function other than to proxy for SES in modelling fertility.

The next step follows from equation 2 whereby the ρ vector is calculated. The elements of this vector are given in table 6, along with the each proxy's correlation coefficient with the dependent variable. Each coefficient ρ_i can be interpreted as a structural regression coefficient in the regression of the particular proxy on the latent variable (permanent income), rescaled so that one unit of permanent income would lead to one additional unit of wave 1 log income. It comes as no surprise that an increase in permanent income leads to an increase in expenditure, wave 3 log income and the number of rooms per person. Additionally, higher SES has a large impact on the probability of owning a computer, microwave, car and landline and leads to a decrease in the probability of owning a bicycle. The relatively low increase in the probability of owning a motorbike can be attributable to the fact that only 1.16% of young adult households have this asset.

The final step is to run the regression and calculate $\hat{\beta}_{LW}$[7]. The next 6 columns report results when all or some of the variables are entered separately into a regression that includes the control and education variables. The first proxy set contains all the proxies and as discussed above only four of the nineteen asset and housing quality variables are statistically significant after controlling for expenditure and income (and therefore SES). In the context of the LW procedure, this is a good result as it means that there are other proxies doing a better job at explaining the dependent variable.

I note than in the case of some of the variables such as radio, television and flushing toilet, their signs are opposite to their raw correlation coefficients. This is not unusual as Lubotsky and Wittenberg (2006) also illustrate that if there are better measured proxies, then the signs of the poorer proxies may be opposite to the true effect of the latent socio-economic variable. Nevertheless their inclusion in the regression provides useful information. The important qualification is that since all the variables do act as proxies for SES, "it is not the individual contributions that matter, but [rather] the aggregate one" (Lubotsky & Wittenberg, 2006, p. 555).

Table 6 - *LW Regression Results based on the 2006 Cape Area Panel Study (CAPS), Sample size: N=1630*

Proxy	Correlation with Birth Ever	ρ Vector	*Proxy Set* 1 Coef.	1 S.E.	2 Coef.	2 S.E.	3 Coef.	3 S.E.	4 Coef.	4 S.E.	5 Coef.	5 S.E.	6 Coef.	6 S.E.
Income (wave 1)	-0.117	1.000	-0.008	(0.014)	-0.050	(0.014)**	-0.020	(0.014)						
Income (wave 3)	-0.091	0.734	0.022	(0.015)	-0.021	(0.015)	0.017	(0.015)						
Expenditure	-0.197	1.441	-0.094	(0.020)**			-0.127	(0.018)**						
Radio	-0.045	0.120	0.027	(0.040)					0.021	(0.040)			0.029	(0.040)
Television	-0.046	0.125	0.090	(0.044)*					0.065	(0.043)			0.084	(0.045)
Video/DVD Player	-0.089	0.362	-0.017	(0.033)					-0.033	(0.033)			-0.022	(0.033)
Landline	-0.101	0.416	-0.012	(0.030)					-0.012	(0.030)			-0.015	(0.030)
Cellphone	-0.111	0.350	-0.062	(0.034)					-0.077	(0.034)*			-0.074	(0.033)*
Refrigerator	-0.092	0.277	-0.041	(0.042)					-0.036	(0.042)			-0.039	(0.043)
Stove	-0.066	0.162	0.015	(0.046)					-0.008	(0.045)			0.007	(0.047)
Microwave	-0.116	0.474	-0.016	(0.032)					-0.040	(0.033)			-0.028	(0.032)
Washing Machine	-0.094	0.390	-0.026	(0.038)					-0.031	(0.039)			-0.037	(0.038)
Bicycle	0.005	-0.014	0.072	(0.032)*					0.074	(0.032)*			0.070	(0.032)*
Motorbike	-0.060	0.054	-0.080	(0.076)					-0.093	(0.076)			-0.113	(0.075)
Car	-0.109	0.402	-0.029	(0.034)					-0.052	(0.034)			-0.042	(0.034)
Computer	-0.159	0.482	-0.045	(0.039)					-0.097	(0.038)**			-0.068	(0.038)
More Than 5 Books	-0.095	0.387	0.042	(0.026)					0.032	(0.027)			0.036	(0.026)
Brick-Type House	-0.079	0.286	0.025	(0.033)							0.010	(0.032)	0.036	(0.033)
Piped Water	-0.077	0.194	-0.075	(0.066)							-0.045	(0.066)	-0.076	(0.067)
Flushing Toilet	-0.074	0.192	0.063	(0.067)							0.029	(0.067)	0.059	(0.068)
Electricity	-0.082	0.134	-0.137	(0.070)*							-0.110	(0.066)	-0.132	(0.071)
Rooms Per Person	-0.185	0.856	-0.071	(0.022)**							-0.139	(0.019)**	-0.118	(0.020)**
Number of Proxies			21		2		3		14		5		19	
LW Estimate			-0.26947951		-0.06566356		-0.18999483		-0.13694441		-0.13395217		-0.21766543	

Note: Robust standard errors in parentheses. * significant at 5%; ** significant at 1%

The dependent variable is an indicator that the woman has had a birth ever. The models include control and education variables.

8.2 – Comparing the LW Estimates

Table 7 displays all the LW estimates of SES as well as the rescaled proxy coefficients from the previous section[8]. I omit the DHS-set estimates as none of them performed as well as their full-set counterparts. This rescaling enables one to compare the estimates. I also report the percentage difference from the LW coefficient for the model including all the proxies as well as the rank of each coefficient.

Table 7 - ***Comparison of LW Estimates based on the 2006 Cape Area Panel Study (2006)***

Proxy Set or Proxy	LW Estimate or Rescaled Proxy Coefficient	Percentage Difference from the Highest	Rank
1 - All	-0.2695	-	1
2 - Incomes Only	-0.0657	75.63%	11
3 - Incomes and Expenditure	-0.1900	29.50%	3
4 - Assets Only	-0.1369	49.18%	5
5 - HQ Only	-0.1340	50.29%	6
6 - Assets and HQ	-0.2177	19.23%	2
Income (wave 1)	-0.0574	78.70%	13
Income (wave 3)	-0.0288	89.30%	14
Average Income	-0.0621	76.95%	12
Expenditure	-0.1833	31.97%	4
SS	-0.0830	69.21%	9
PCA	-0.0813	69.81%	10
SS + HQ	-0.1132	57.98%	7
PCA + HQ	-0.0862	68.02%	8

Note: All LW estimates and proxy coefficients are on the same scale.

As I have noted in the previous section, the simple average income proxy performs better than either of the observed individual wave income proxies. However the estimate on the regression weighted average (proxy set 2) is even less attenuated. Once expenditure is added (proxy set 3), the attenuation drops substantially. When comparing this estimate to that of expenditure alone, we observe that there is not much marked difference indicating that the income proxies are of little *additional* value when using the expenditure proxy.

When all 21 proxies are used (proxy set 1), the estimated effect of permanent income is -0.2695. Whereas that effect when using only the strongest proxy, expenditure, is approximately 32% lower. This clearly shows that using assets to augment the information from a proxy, even a relatively well measured proxy like expenditure, is beneficial.

The coefficient on the regression weighted index of the full set of assets (proxy set 4) is over 60% larger than each of the adjusted coefficients on the PCA and SS indices. Interestingly enough, the regression weighted index of only the housing quality measures (proxy set 5) performs almost as well as that of only the assets (proxy set 4). In addition, both sets also perform better than those PCA and SS indices that include both the assets and the housing quality measures. In explaining the effect of SES, the LW procedure is therefore less sensitive to the number of assets used when compared to the PCA and SS.

Not surprisingly, the LW estimate when both the assets and housing quality measures are included (proxy set 6) is even less attenuated. In fact, the absolute value of this coefficient is higher than that of the model that only has the income and expenditure proxies (proxy set 3). This signifies that in the absence of income or expenditure data, using assets and housing quality measures to control for SES in fertility analysis is a worthwhile endeavour despite the potential lack of variability that may exist across households.

8.3- Effect of the Covariates

The fertility model used in my empirical work includes a set of controls and education variables. In contrast to the 'mis-measured' proxies, these covariates are correctly measured. Lubotsky and Wittenberg (2006) show that if any of these correctly measured covariates are related to any of the proxy variables through the measurement error component, then the method of moments estimator rho becomes inconsistent. In addition, the bias on the covariates resulting from the measurement error would be exacerbated. The authors offer a covariates-adjusted estimate that uses the residuals from a regression of the proxies and dependent variable on the covariates, rather than the actual variables.[9] Although the models include age, race, never married and education covariates, I follow Lubotsky and Wittenberg (2006) and refrain from using the covariates-adjusted estimator in this comparative study, despite the presence of covariates.

9 – CONCLUSIONS

It is evident that the relationship between SES and fertility is multi-dimensional and interwoven with a plethora of peripheral and intermediate variables. Apart from the lack of consensus on the underlying mechanism behind the effect of SES on fertility, there is no agreement on which measures or methods most reliably capture this effect. Each proxy has its own set of limitations, but nevertheless, offers valuable information.

The purpose of this paper is two-fold. In a similar vein to Bollen, Glanville, and Stecklov (2002), I wanted to first learn whether the particular measures and methods used to incorporate SES proxies into fertility studies actually make a difference. My research covers a variety of proxies currently employed, and determines their impact on analyses. I secondly aimed at utilising the wealth of information available in the diverse set of proxy variables to better understand the relationship between SES and childbearing.

With regards to the first aspect, I sought to identify if the choice of proxy influenced the inferences made about the relationship between SES and fertility, as well as that between other explanatory variables and fertility. From my analysis of the CAPS survey data from

South Africa; I conclude that the measure and method used to capture the effect of SES does matter. Different proxies lead to different estimated impacts of both SES and certain other explanatory variables.

The impact of maternal education changes considerably from one proxy to the next, while that of age is more constant. In addition, maternal education was found to exhibit both direct effects as well as those mediated through SES. While all the SES proxies have a negative and significant effect on fertility, consumption expenditure has the largest standardised and un-standardised coefficients, as well as the highest R-squared value indicating the best relative fit. In contrast, certain restricted DHS-set asset indices preformed remarkably poorly. Thus, if a researcher's interest lies in the effects of education or SES, he or she must therefore approach the measures with caution as the choice makes a difference in explaining childbearing outcomes.

In dealing with an abstraction like SES, one cannot neglect the issue of measurement error and the associated potential problems arising from the use of proxies. However, the surge of research into the use of consumer durable goods and housing quality measures to create asset indices has opened a window for opportunity in the field of fertility. In attempting to better estimate the fertility effect of SES, I use the LW procedure to counter some of the attenuating effects of measurement error by extracting as much information from all the proxy variables as possible. The LW estimation procedure "cannot in any real way 'fix' bad data" but its regression weighted index is superior in that it produces an estimate with a minimized attenuation bias (Wittenberg, 2005, p. 23).

Overall, I conclude that SES is latent and in the case of the childbearing behaviour of *young* women, it is their socio-economic background that is the key operator. Having more detailed information about household consumer durable goods and housing quality measures can be extremely useful, especially if they do in fact proxy for the latent socio-economic variable. These asset variables can be used to create a regression weighted index that can augment the information from other proxies (regardless of the degree of measurement error) as well as 'substitute' for otherwise absent income or expenditure data, improving the estimated effect of SES in both cases.

In order for the LW estimates of SES to be used with greater precision, more research needs to go into uncovering any possible direct effects that asset variables may have with fertility. Wittenberg (2007) provides a specification test that helps identify structural relationships and if used in conjunction with the more efficient covariates-adjusted LW estimator, a more accurate socio-economic effect can be estimated.

APPENDIX

A1 - Imputations

Only 2 out of the 1630 observations from wave 1 had missing income data as opposed to 239 (14.66%) from wave 3. I therefore imputed the median income value by race (population group).

The total household expenditure variable comprises of individual expenditure items such as food, rent, clothing and transport. Consequently the number of missing values varied across the expenditure items ranging from 8 to 203. I therefore imputed the median expenditure item value by race (population group) before aggregating and dividing by household size.

A2 – Principal Components Analysis

Table A2 - ***Principal Component Analysis Scoring Coefficients for the 2006 Cape Area Panel Study (CAPS)***

	PCA Scores	PCA + HQ Scores	DHS-PCA Scores	DHS-PCA + HQ Scores
Own Radio	0.2109	0.1832	0.4351	0.2447
Own Television	0.2773	0.2561	0.5158	0.3466
Own Video/DVD Player	0.3359	0.2828		0.3704
Own Landline	0.3175	0.2710		
Own Cellphone	0.1179	0.0871		
Own Refrigerator	0.2993	0.2788	0.5262	
Own Stove	0.2368	0.2241		
Own Microwave	0.3498	0.2927		
Own Washing Machine	0.3587	0.3056		
Own Bicycle	0.1966	0.1494		
Own Motorbike	0.0765	0.0570	0.2949	0.1403
Own Car	0.2940	0.2355	0.1408	0.0582
Own Computer	0.2670	0.2102	0.4010	0.2282
Own More Than 5 Books	0.2284	0.1874		
Brick-Type House		0.2819		0.3828
Piped Water		0.2533		0.4045
Flushing Toilet		0.2552		0.4085
Electricity		0.2029		0.3318
Rooms Per Person		0.1504		0.1434
Eigenvalue	4.6279	5.8547	0.0594	3.6315
Proportion of variation captured	0.3306	0.3081	0.3432	0.3301

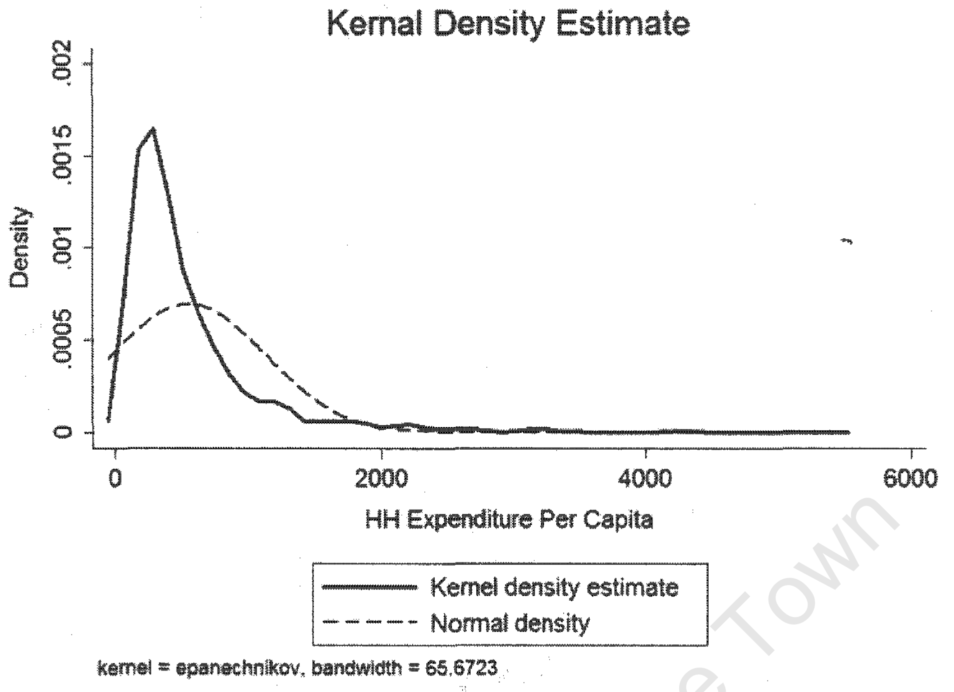
A3 – Descriptive Statistics

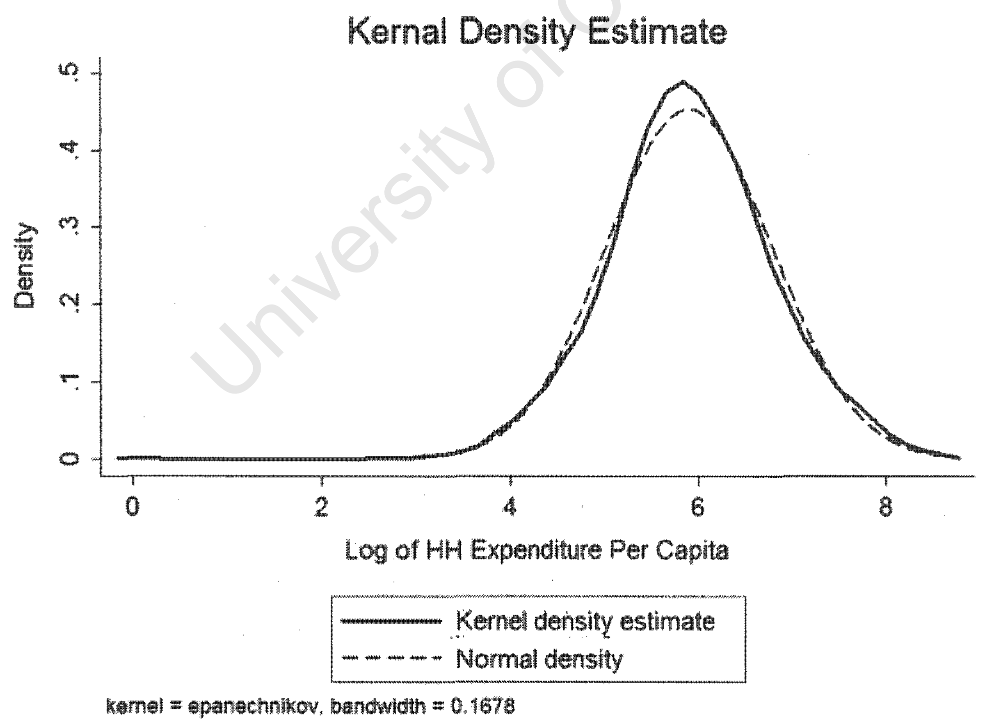
Table A3 - ***Original Variable Descriptive Statistics in the 2006 Cape Area Panel Study (CAPS)***

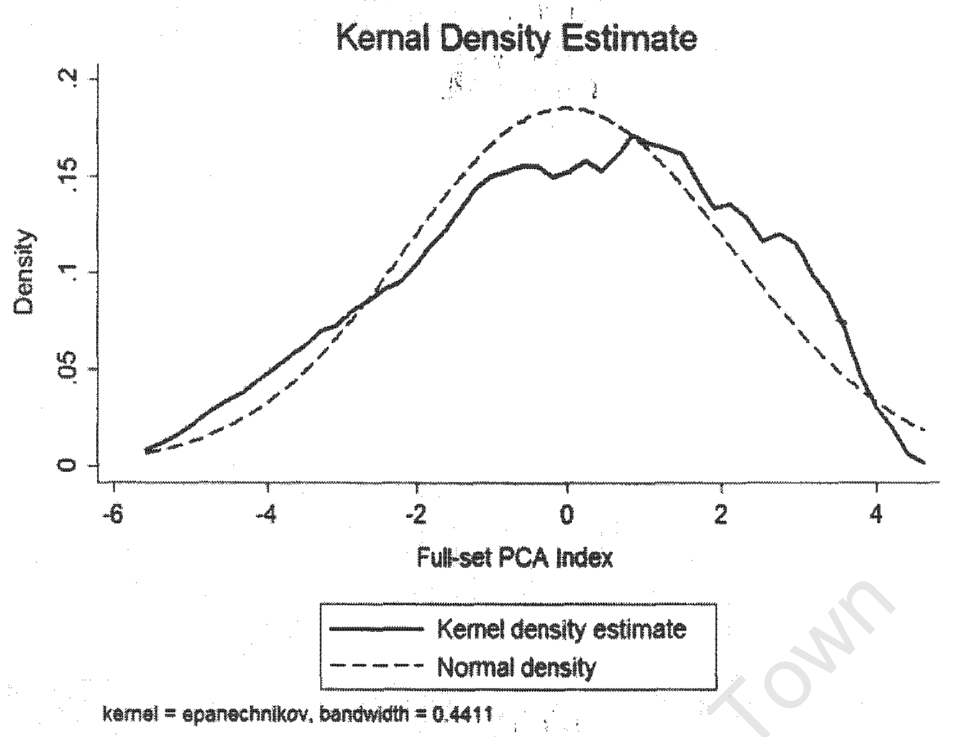
Original Variables	Mean	Std. Dev.
Un-Logged Proxies		
Per Capita Income wave 1	626.9153	749.5641
Per Capita Income wave 3	790.5601	910.1681
Expenditure Per Person	552.9029	575.7535
Asset Indicators		
Own Radio	**0.8828221**	0.3217309
Own Television	**0.8779141**	0.3274855
Own Video/DVD Player	0.5920245	0.4916094
Own Landline	0.4343558	0.4958242
Own Cellphone	**0.8263804**	0.3788983
Own Refrigerator	**0.8447853**	0.3622204
Own Stove	**0.9042945**	0.2942773
Own Microwave	0.5889571	0.4921740
Own Washing Machine	0.5190184	0.4997915
Own Bicycle	**0.1503067**	0.3574815
Own Motorbike	**0.0116564**	0.1073669
Own Car	**0.2717791**	0.4450132
Own Computer	**0.1576687**	0.3645419
Own More Than 5 Books	0.5883436	0.4922846
Housing Quality Measures		
Brick-Type House	0.7435583	0.4368025
Piped Water	**0.8987730**	0.3017216
Flushing Toilet	**0.8914110**	0.3112183
Electricity	**0.9601227**	0.1957310
Rooms Per Person	0.8757077	0.5579957
N=1630		

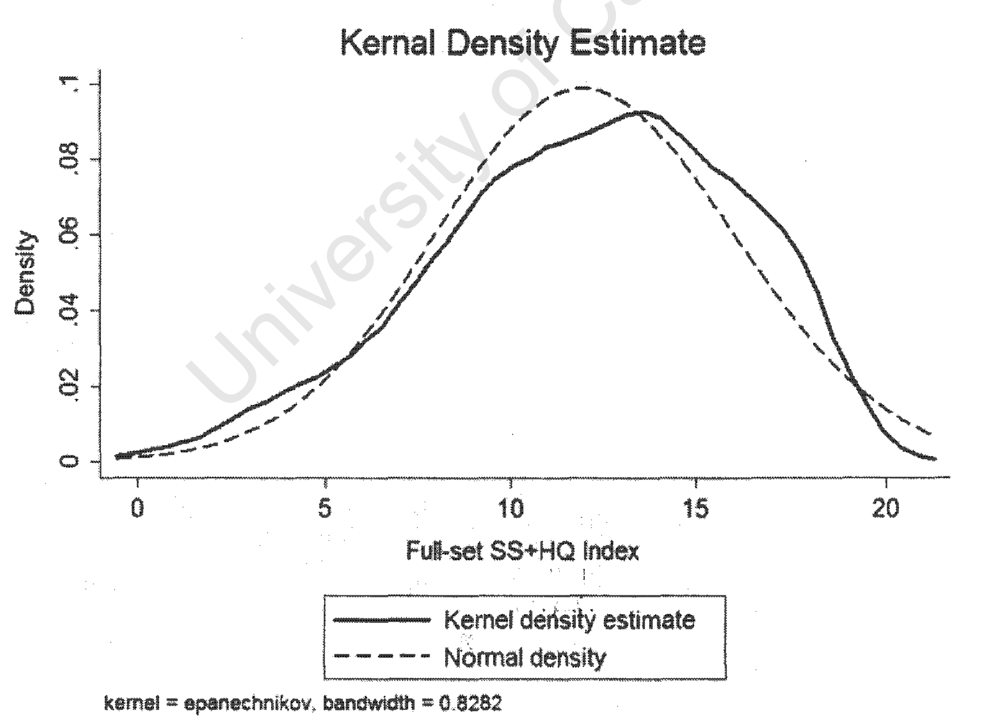
Note: **bold** indicates mean values of indicator variables greater than 0.8 or lower than 0.2

A4 – Kernel Density Estimates






Kernal Density Estimate
Density
0
.05
.1
.15
.2
-6
-4
-2
0
2
4
Full-set PCA Index
Kernel density estimate
Normal density
kernel = epanechnikov, bandwidth = 0.4411


Kernal Density Estimate
Density
0
.02
.04
.06
.08
.1
0
5
10
15
20
Full-set SS+HQ Index
Kernel density estimate
Normal density
kernel = epanechnikov, bandwidth = 0.8282

A5 – Auxiliary Analysis

I conducted supplementary regressions to examine whether the 'background effect' hypothesis could be true. Since over 85 percent of the young women are neither the heads of the households nor wives, partners or girlfriends to the heads of the households, they do not assume the role of the female household head in the typical nuclear family. It stands to reason that if the proxies are picking up on the signal from the socio-economic background of the vast majority of the young women, then this signal is being distorted by the current SES of the minority of women who do fulfil the role of the female household head. One would then expect an increase in the estimated effects of the proxies if one excludes that minority of women (or households) from the sample.

Table A5 reports two sets of coefficients and a set of percentage differences. The 'full sample' refers to the current sample under investigation in this paper. The 'non-nuclear family' subsample, of course, refers to the full sample without the above mentioned minority.

Table A5 - ***Percentage Difference Between the Coefficients of the Full Sample and the 'Non-Nuclear Family' Subsample based on the 2006 Cape Area Panel Study (CAPS)***

Proxy	'Full Sample' Coefficient	'Non-Nuclear Family' Subsample Coefficient	Percentage Difference (Subfamily-Full Sample)/Full Sample
Income (wave 1)	-0.0573858	-0.0589091	**2.65%**
Income (wave 3)	-0.0392777	-0.0374611	-4.63%
Average Income	-0.0716204	-0.070693	-1.29%
Expenditure	-0.1272534	-0.1319481	**3.69%**
SS	-0.0208064	-0.0213768	**2.74%**
PCA	-0.0311342	-0.0319715	**2.69%**
SS + HQ	-0.0200448	-0.0200119	-0.16%
PCA + HQ	-0.0280336	-0.0284595	**1.52%**
DHS-SS	-0.0269568	-0.0256401	-4.88%
DHS-PCA	-0.0203106	-0.0179467	-11.64%
DHS-SS + HQ	-0.0300493	-0.0287406	-4.36%
DHS-PCA + HQ	-0.0189642	-0.0160691	-15.27%

Note: **bold** indicates positive differences

Indeed the coefficients on almost all the 'well performing' proxies increase. The coefficient on expenditure shows the largest percentage increase when compared to any of the other proxies. The wave 1 observed income proxy, already a good measure of socio-economic background as explained in section 7.3, also has a higher effect. Of the full-set asset indices, three of the four show marked increases in performance. Only the impact of the simple-sum including housing quality marginally decreases. Not surprisingly, the poorly performing wave 3 observed income and DHS-set proxies perform even worse with a smaller sample.

BIBLIOGRAPHY

Balk, D. (1994). Individual and Community Aspects of Women's Status and Fertility in Rural Bangledesh. *Population Studies* , *48* (1), 21-45.

Becker, G. S. (1981). *A Treatise on the Family.* Cambridge, MA: Harvard University Press.

Bollen, K. A., Glanville, J. L., & Stecklov, G. (2002). Economic Status Proxies in Studies of Fertility in Developing Countries: Does the Measure Matter. *Population Studies* , *56* (1), 81-96.

Bollen, K. A., Glanville, J. L., & Stecklov, G. (2001). Socioeconomic Status And Class In Studies of Fertility And Child Health In Developing Countries. *Annual Review Of Sociology* , *27*, 153-185.

Bollen, K. A., Glanville, J. L., & Stecklov, G. (2007). Socio-economic Status, Permanent Income, and Fertility: A Latent-variable Approach. *Population Studies* , *61* (1), 15-34.

Bollen, K. A., Guilkey, D. K., & Mroz, T. A. (1995). Binary Outcomes and Endogenous Explanatory Variables - Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia. *Demography* , *32* (1), 111-131.

Booysen, F., Van Der Berg, S., Burger, R., Von Maltitz, M., & Du Rand, G. (2008). Using an Asset Index to Assess Trends in Poverty in Seven Sub-Saharan African Countries. *World Development* , *36* (6), 1113-1130.

Bouis, H. E. (1994). The Effect of Income on Demand for Food in Poor Countries: Are our Food Consumption Databases Giving us Reliable Estimates? *Journal of Development Economics* , *44* (1), 199-226.

Cortinovis, I., Vella, V., & Ndiku, J. (1993). Construction of a Socio-economic Index to Facilitate Analysis of Health Data in Developing Countries. *Social Science and Medicine* , *36* (8), 1087-1097.

Dargent-Molina, P., James, S. A., Strogatz, D. S., & Savitz, D. A. (1994). Association between Maternal Education and Infant Diarrhea in Different Household and Community Environments. *Social Science and Medicine* , *38* (2), 343-350.

Davis, K., & Blake, J. (1956). Social Structure and Fertility: an Analytical Framework. *Economic Development and Cultural Change* , *4* (3), 211-235.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore and London: The John Hopkins University Press.

Desai, S., & Alva, S. (1998). Maternal Education and Child Health: Is There a Strong Causal Relationship? *Demography* , *35* (1), 71-81.

Devin, R. B., & Erickson, P. I. (1996). The Influence of Male Care Givers on Child Health in Rural Haiti. *Social Science and Medicine* , *43* (2), 479-488.

Easterlin, R. A. (1969). Towards a Socioeconomic Theory of Fertility. In S. J. Behrman, L. Corsa, & R. Freedman, *Fertility and Family Planning: A World View* (pp. 127-156). Ann Arbor: University of Michigan Press.

Falkingham, J., & Namazie, C. (2002). Measuring Health and Poverty: A Review of Approaches to Identifying the Poor. London: Department for International Development Health Systems Resource

Centre (DFID HSRC). Retrieved January 27, 2009, from http://www.eldis.org/go/display/?id=12061&type=Document

Filmer, D., & Pritchett, L. (2001). Estimating Wealth Effects Without Income Or Expenditure Data - Or Tears: Educational Enrollment In India. *Demography*, *38* (1), 115-132.

Filmer, D., & Pritchett, L. (1999). The Effect Of Household Wealth On Educational Attainment: Evidence From 35 Countries. *Population And Development Review*, *25* (1), 85-120.

Friedman, M. (1957). *A Theory Of The Consumption Function.* Princeton: Princton University Press.

Gorbach, P. M., Hoa, D. T., Nhan, V. Q., & Tsui, A. O. (1998). Contraception and Abortion in Two Vietnamese Communes. *American Journal of Public Health*, *88* (4), 660-663.

Gujarati, D. N. (2003). *Basic Econometrics* (4th ed.). McGraw Hill.

Hentschel, J., & Lanjouw, P. (1996). Constructing An Indicator Of Consumption For The Analysis Of Poverty: Principles And Illustrations With Reference To Ecuador. *Living Standards Measurement Study*, *Working Paper* (124).

Hoddinott, J., & Haddad, L. (1995). Does Female Income Share Influence Household Expenditures? Evidence From Cote D'Ivoire. *Oxford Bulletin of Economics and Statistics*, *57* (1), 77-96.

Houweling, T. A., Kunst, A. E., & Mackenbach, J. P. (2003). Measuring Health Inequality Among Children in Developing Countries: Does the Choice of the Indicator of Economic Status Matter? *International Journal for Equity in Health*, *2* (8).

Howe, L. D., Hargreaves, J. R., & Huttly, S. R. (2008). Issues in the Construction of Wealth Indices for the Measurement of Socio-economic Position in Low-Income Countries. *Emerging Themes in Epidemiology*, *5* (3), doi:10.1186/1742-7622-5-3.

Lam, Lam, D., Ardington, C., Branson, N., Case, A., Menendez, A., et al. (August 2008). *A Very Short Introduction to the Cape Area Panel Study Integrated Waves 1-2-3-4 Data.* University of Cape Town.

Lubotsky, D., & Wittenberg, M. (2006). Interpretation of Regressions with Multiple Proxies. *The Review of Economics and Statistics*, *88* (3), 549-562.

McKenzie, D. J. (2005). Measuring Inequality with Asset Indicators. *Journal of Population Economics*, *18* (2), 229-260.

Montgomery, M. R., Gragnolati, M., Burke, K. A., & Paredes, E. (2000). Measuring Living Standards With Proxy Variables. *Demography*, , *37* (2), 155-174.

Moultrie, T. A., & McGrath, N. (2007). Teenage fertility rates falling in South Africa. *South African Medical Journal*, *97* (6), 442-443.

Muhuri, P. K. (1996). Estimating Seasonality Effects on Child Mortality in Matlab, Bangladesh. *Demography*, *33* (1), 98-110.

Raftery, A. E. (1995). Bayesian Selection in Social Research. *Sociological Methodology*, *25*, 111-163.

Raftery, A. E., Lewis, S. M., & Aghajanian, A. (1995). Demand or Ideation? Evidence from the Iranian Marital Fertility Decline. *Demography*, *32* (2), 159-182.

Robinson, W. C. (1997). The Economic Theory of Fertility over Three Decades. *Population Studies*, *51* (1), 63-74.

Schwarz, G. (1978). Estimating The Dimension Of A Model. *Annals of Statistics*, *6* (4), 461-464.

Solon, G. (1992). Intergenerational Income Mobility in the United States. *American Economic Reveiw*, *82* (3), 393-408.

Vaya, S., & Kumaranayake, L. (2006). Constructing Socio-economic Status Indices: How to Use Principal Component Analysis. *Health Policy and Planning*, *21* (6), 459—468.

Wittenberg, M. (2007, March). Testing for a Common Latent Variable in a Linear Regression. School of Economics, University of Cape Town. Available at http://mpra.ub.uni-muenchen.de/2550/1/MPRA_paper_2550.pdf.

Wittenberg, M. (2005, October). Testing for a Common Latent Variable in a Linear Regression: Or How to "Fix" a Bad Variable by Adding Multiple Proxies for it. CSSR Working Paper No. 132. Available at http://www.saldru.uct.ac.za/papers/cssrwps/wp132.pdf.

Wittenberg, M. (forthcoming). Weighing the Value of Asset Proxies: The Case of the Body Mass Index in South Africa. Working Paper, Datafirst/SALDRU, University of Cape Town.

NOTES

[1] SES can only influence fertility by operating through a set of intermediate or proximate determinants such as age at marriage, contraceptive use or involuntary infecundity. These determinants directly affect an individual's exposure to intercourse, conception and gestation (Davis & Blake, 1956). Reduced-form implies that all such proximate influences are ignored, thus accounting only for the exogenous total effect of SES on fertility.

[2] While often used synonymously, 'method' and 'methodology' are two distinct terms. The latter pertains to the theory or systematic study of methods used within a discipline. For a discussion of the methodological issues regarding the treatment of SES in certain behavioural outcomes I refer the interested reader to Bollen, Glanville and Stecklov (2001).

[3] See the CAPS website for free access to the data: http://www.caps.uct.ac.za/
Income data from wave 4 was unavailable therefore income data from waves 1 and 3 were used. Apart from these two variables, all data was obtained from the most recent wave, wave 4 (2006).

[4] Both logit and probit regressions were run in addition to the linear probability models to test for any substantial differences. The results were in line with those of the linear probability models in terms of signs and significant variables. Of the three, the linear probability model was the preferred choice of estimation primarily because its coefficients are the easiest to interpret.

[5] The graphs in appendix A4 indicate that the distributions underlying the majority of the proxies are roughly normal. This strengthens the justification for using standardised coefficients as a means of comparison. A standard deviation increase in the one proxy can therefore be viewed as being equivalent to a standard deviation increase in another.

[6] Wittenberg (forthcoming) shows that only in the case for binary variables, does an asset that is more unequally distributed be assigned a heavier weight. Continuous variables (such as number of rooms per person) with low standard deviations will actually receive a higher weight due to the division of the standard deviation inherent in PCA.

[7] Although calculation of $\hat{\beta}_{LW}$ is rather simple, Lubotsky and Wittenberg (2006) recommend bootstrapping methods for calculations of the standard errors. However these standard errors are of little relevance in this study so I refrain from reporting any.

[8] In order to place the original coefficients on the same scale as the LW estimates, each coefficient must be multiplied by $\rho'\delta$, where ρ is the vector in column 2 of table 6 and δ is the vector of weights for the linear combination of proxies in question. In the case of the simple average income proxy, the weight on each income variable is given by $\delta_i = 0.5$, therefore $\rho'\delta = {(1 + \rho_2)}/{2}$. In a PCA $\delta_i = {score_i}/{\sigma_i}$ where $score_i$ is the scoring factor (eigenvector) for variable i and σ_i is the standard deviation of that variable. Each proxy in a simple-sum is given an equal weighting of one. Technically, any individual proxy used on its own can be written as linear combination of any number of hypothetical proxies. In this case the weight assigned to the individual proxy used is one and that of every other hypothetical proxy is zero which means that for individual variable i, $\rho'\delta = \rho_i$.

[9] Wittenberg (2007) proposes a more efficient version of this covariates-adjusted estimator.