Journal of African Economies, Volume 12, Number 3, pp. 371–416

Testing for Son Preference in South Africa

Lata Gangadharan[a] and Pushkar Maitra[b1]
[a]University of Melbourne and [b]Monash University

Evidence from many developing countries suggests that parents have a preference for sons over daughters. This has been referred to as son preference. This paper uses individual level unit record data to test the son preference hypothesis in South Africa. We use an accelerated hazard model to estimate the duration between successive births and our results indicate that son preference exists only for the Indian community in South Africa. Indian households are observed to have a higher duration between children following the birth of a son, irrespective of the number of children they already have. For the rest of the population, there is very little evidence of son preference. Preference for sons could be the result of a combination of factors including religious beliefs and social customs such as the dowry system, lineage and familial and kinship ties.

1. Introduction

The gender preference hypothesis postulates that parents exhibit preferences for having children of a particular gender. In many developing countries, parents seem to have a preference for sons over daughters. Such preferences could be the result of any combination of social, cultural and economic factors — for example, in most developing countries, sons continue to stay at home, augment household

[1] We would like to thank Siwan Anderson, Bob Bartels, Monojit Chatterji, Patrick Francois, Brett Inder, Elizabeth Savage, Don Wright, two anonymous referees, seminar participants at the University of Sydney, participants at the 8th Econometric Society World Congress held at Seattle, participants at the 14th European Society for Population Economics Annual Meetings held at Bonn and participants at the 28th Australian Conference of Economists held at La Trobe University for comments and discussions. We are, however, responsible for all remaining errors. Funding was provided by the Faculty Research Grant Scheme, Faculty of Economics and Commerce, University of Melbourne and the Faculty Research Grant Scheme, Faculty of Business and Economics, Monash University.

income and provide old-age parental support. Daughters, on the other hand, marry and move to another household. Son preference has important social and economic implications and can substantially influence patterns of fertility, child mortality and intra-household allocation of resources.

There now exists a large literature that documents son preference among parents in different developing countries (though primarily Asian countries). Evidence on son preference is widespread in the Indian subcontinent — see Sen and Sengupta (1983), Dasgupta (1987), Kishor (1993) and Arnold *et al*. (1998) for evidence from India, Ali Khan and Sirageldin (1977) and Gangadharan and Maitra (2000) for evidence from Pakistan and Chowdhury and Bairagi (1990) and Rahman and Da Vanzo (1993) for evidence from Bangladesh. Evidence shows that son preference affects the actual childbearing intentions of parents and results in discrimination against girls. Such discrimination is often reflected in the form of extreme malnutrition of girls and also significantly higher child mortality rates among girls.[2] In African countries (both in South Africa and elsewhere), the evidence is mixed. Research using data from Ghana (Garg and Morduch, 1998; Morduch, 2000) finds that while there is a strong positive association between educational and health outcomes and the number of sisters a particular child has (holding constant the number of siblings), this association is not affected by the gender of the child. On the other hand, Aly and Shields (1991) show that in Egypt the probability of a woman using contraception increases with the number of existing male children. Thomas (1994), using data from Ghana, finds that parental resources have differential impacts depending on the gender of the child, with mothers' resources having a greater impact on the health outcome of girls and fathers' resources having a greater impact on the health outcome of boys. Turning to evidence from South Africa, Morduch (2000) finds that within Black households there is no association between the number of sisters (holding constant the number of siblings) and schooling outcomes of children. Quisumbing and Maluccio (2000), which to the best of our knowledge is the only other paper that examines whether gender bias is culturally rooted,

[2] In other Asian countries, evidence of son preference is quite widespread as well. For example, Leung (1988) and Raut (1996) find evidence of preference for sons among the Chinese but not the Malays in Malaysia. Larsen *et al*. (1998), using data from Korea, find that women who have a son are less likely to have another child and if they do have another child, the duration between children is longer.

find that in Indian households men and women have significant differences in gender preferences. Using data from the Kwazulu-Natal province in South Africa, they find that Indian women significantly favour their daughters, while Indian men significantly favour their sons.

This paper re-examines the son preference hypothesis in South Africa. The variable of interest in our analysis is the duration between successive births. The duration between successive births is important because household resources are limited and a shorter duration between births increases the competition between siblings for finite household resources, leading to a decline in child quality. If there is indeed a preference for sons over daughters, we expect to observe the following: first, an increased duration following the birth of a son and, secondly, the higher the number of existing sons, the greater the duration between successive children. To the best of our knowledge, this is the first paper that uses duration between successive births to measure son preference in Africa. The existing literature uses other manifestations of son preference. For example, Morduch (2000) and Quisumbing and Maluccio (2000) use educational outcomes of boys and girls to test for evidence of son preference.

South Africa provides an interesting alternative to the Asian countries and is also of interest because of the sharp racial divide between its various communities.[3] Moreover, the presence of a sizeable Indian community in South Africa sets the stage for an interesting comparison with countries in Asia. South Africa ranks as an upper-middle income country, with a per-capita GDP of $3,000 (Carter and May, 1999) and the Indian community in South Africa forms the core of a rich, urban and professional upper-middle class, with an average monthly household income of rand (R) 3,347.73 (nearly $1,000) in 1993–4.

Most Indians live in Natal (in 1980, around 73% of all South African Indians lived in the Durban–Pinetown area, where they constituted around 37.5% of the population) and restrictions on movement and property ownership, distance from Natal and the presence of strong community institutions and family ties have discouraged movement

[3] During the apartheid era, all South Africans were classified into one of the four race categories: Black (African), Coloured (mixed race), Indian (Asian) and White (Caucasian). To maintain consistency with the data and the existing literature, we use this categorisation in our paper.

of Indians out of Natal.[4] The majority of Indians are Hindus (in 1993, more than 60% of all Indians were Hindus) and the Whites have always viewed Indians as being 'alien and inassimilable because of visible views of pluralism like religion, food and dress' (Lemon, 1987). The extended family (*kutum*) is a very important aspect of Indian life in South Africa and it is this extended family that has, over time, insulated Indians from the oppressions of the South African society under apartheid. The Indian *kutum* is highly male dominated and consists of all families who can be traced to a common paternal grandfather or the grandfather's brother and, over time, has come to extend across barriers of religion, language, economics and caste. Under such circumstances, it is not surprising that Indians portray a strong preference for sons. While it is true that the extended family and paternal society make Indians different from the rest of the South African population, it is, however, surprising that even though the Indians in South Africa have a history going back to 1860s, they still seem to behave very similarly to Indians who have always lived in the Indian sub-continent, have rural backgrounds and come from low income communities.[5] The Indian immigration to South Africa has been limited to the dependents of existing residents in 1913, when the importation of indentured labour ceased and by 1960 only 5.5% of Indians were born outside of South Africa. Of the 890,000 South African Indians, more than 70% are descendants of the indentured immigrants and return migration to India was not allowed. Today, Indians in South Africa form a rich, urban upper middle-class, but their preferences regarding the gender of their children are very similar to Indians in the sub-continent. See Kuper (1960) and Lemon (1987) for excellent surveys of Indians in South Africa. The extended family system is quite important amongst the Blacks as well. Child fostering is common among Blacks in South Africa and there is evidence that adult males (and females) would leave their children with the grandparents in the villages and migrate to the cities to work. This is referred to as the 'oscillatory migratory system'. Many Blacks

[4] Indians were not allowed to travel through Orange Free State and they were not allowed to own property in Transvaal. While there were no restrictions on property ownership in Cape Province, very few Indians moved there because of the distance from Natal.

[5] Indian society is termed as being paternal in the sense that the extended family is linked to the common paternal grandfather.

use this system to retain their cultural and tribal ties to the rural areas (Jooma, 1991).

Our results show that the son preference hypothesis is supported only for the Indian households. For the Indian households, the duration between children is higher after the birth of a son. Additionally, an increase in the number of existing sons increases the duration between the second and third children. For the rest of the population, there is no evidence of son preference. An anonymous referee enquired whether this represents 'behaviour favouring sons' as opposed to 'preference for sons'. While there is an important difference between these two terms, often it is not easy to distinguish between them econometrically. 'Behaviour favouring sons' typically arises because of economic reasons. Faced with limited resources, parents may often favour sons because the expected return from sons is greater than the expected return from daughters. Rosenzweig and Schultz (1982) argue that children who are expected to be economically more productive as adults receive a larger share of family resources and have a higher quality of life (are healthier, earn more or are better educated). In most developing countries, male children typically stay with their parents even as adults and contribute to the household income, while daughters move to their husband's home once married. Therefore, the expected household income is higher if the child is a son. 'Preference for sons', on the other hand, may be defined more broadly. Economic reasons of the kind outlined above can result in 'preference for sons', but parents in many societies may also prefer sons because of non-economic reasons. For example, in many societies, norms dictate that sons take care of their elderly parents and it is sons that 'carry on the family name' or 'light the funeral pyre'.

Our paper contributes significantly to the literature on gender preferences as it is, to the best of our knowledge, the first to study son preference among South African households using this particular approach. The presence of four distinct races makes South Africa an interesting and important country to study. The estimation methodology used in this paper is a significant improvement over the existing literature. First, the use of the hazard model allows us to account for censored observations in the sample, arising from the presence of women who have not 'exited' at each transition.[6] Ordinary least

[6] Transition is the duration between successive children. This is explained in detail in section 2.

squares techniques do not allow us to account for censoring in the data. Analyses of duration between children generally use a proportional hazard model, which leads to possible mis-specification because it does not take into account the effect of time on the hazard of having another child. In this paper we use an accelerated hazard analysis of the duration between successive births. That allows us to capture the effect of time, an issue that could be of significant importance in the analysis of duration between births. To the best of our knowledge, the only previous paper that actually takes into account the effect of time on the duration between births is Raut (1996). However, our econometric analysis is an improvement over that of Raut in that we use a gamma distribution to characterise the baseline hazard function, which is the most general distribution and encompasses the log normal and the Weibull distributions used by Raut (1996). The use of the gamma distribution provides us with the flexibility to model a non-monotonic hazard function.

The rest of the paper is organised as follows. Section 2 describes the methodology used in the paper. Section 3 describes the data and selected descriptive statistics. Section 4 discusses the results and Section 5 concludes.

2. Methodology

Let us start with the assumption that the quality Q (education, earnings, health, or any other measure) of a child born to a household depends on its birth order, the age of its parents when born, the interval between its birth and prior and subsequent births, household (parental) endowment, child specific resources and child specific (quality) endowments. Parents are assumed to care about child quality, not only because parents care for the welfare of their children, but also because higher child quality typically implies higher earnings (as adults) and this in turn increases household income. Following Rosenzweig (1986), in this set up one can show that (i) more endowed parents will tend to space births more closely and (ii) if endowment of previous children increases, then the spacing between children will increase.

We use the highest level of education attained by the mother as the measure of parental endowment. This is, however, not a perfect measure and might not fully capture the household (or parental endowment) effect and therefore we examine the robustness of the results using two alternative measures — the highest education attained by the household head and the income of the household. It might, however, be noted that there are problems with the use of both these measures as well, arising primarily from the fact that the data set is not retrospective in nature.

We use the sex of the child as the measure of child endowment. It is seen that parents in many societies regard their male children as having higher endowments compared to their female children. This could be due to both economic and non-economic reasons. For example, sons are regarded as better insurance for the parents' old age as they have a higher probability of staying with their parents and contributing to household income.[7]

The duration between successive births is modelled as a failure-time process represented by a log-hazard of duration equation. Let T be the duration of an event, such as the duration between the first and the second birth, the duration between the second and third birth, etc. Let u be a strategy (such as the use of a contraceptive method) that the woman might adopt to control the duration of the event and where U is the set of all feasible strategies. T will depend on a number of factors, not all of which are observable to the researcher (for example, the biological endowment of the woman). Let η denote the set of all such unobservable factors that we will call individual specific unobserved heterogeneity. Then the hazard rate of an event T can be defined as $h(t \mid u,\eta) \equiv$ probability that the event T occurs in the time interval $(t,t + \mathrm{d}t)$, given that it has not occurred until t and given the value of the individual specific unobserved heterogeneity (η) and the actual strategy followed (u).

Let $\eta = 0$ and $u = 0$ represent a woman with an average level of biological endowments who has not followed any specific strategy for childbirth. Then the baseline hazard function is defined as

$$\lambda_0(t) = h(t \mid u = 0, \eta = 0).$$

The effect of a particular strategy adopted or specific biological endowments is to scale the baseline hazard up or down as follows:

[7] The sex of the child is the only measure of child endowment that we use in this paper. However, there are alternative measures. For example, child intelligence could be a measure of child endowment, as could child health. The problem with using variables such as child intelligence and child health is that information on these variables is private to the parents and is unobserved. The sex of the child, on the other hand, is observable to the researcher.

$$h(t \mid u,\eta) = \lambda_0(t)\Psi(u,\eta),\ \Psi > 0.$$

Let X denote the co-variates whose values represent the information available to the woman at t. The specific strategy adopted will then depend both on X and on the unobserved heterogeneity, so that $u = u(X,\eta)$. If we impose the restriction

$$\Psi(u(X,\eta),\eta) = e^{X'\beta + \eta}$$

then the proportional hazard model for the observed spacing between births can be written as

$$h(t \mid X,\eta) = \lambda_0 e^{X'\beta}.$$

The use of proportional hazard models for the spacing of births could, however, lead to a misleading description of observed choices and, hence, result in incorrect policy prescriptions. In particular, the proportional hazard model does not take into account the effect of time on the hazard function. This causes problems, particularly in analysing issues such as the duration between successive births, where the hazard function is likely to depend on time. To account explicitly for the effect of time, we consider the accelerated hazard model. The proportional hazard model and the accelerated hazard model both estimate the same model but in different metrics — if, for example, the set of coefficients in the proportional hazard model is denoted by β and the coefficients in the accelerated hazard model are denoted by β^*, then $\beta^* = -\beta/\kappa$. A negative coefficient β^* therefore indicates a higher hazard ratio and, hence, a decreased duration between successive children.

The use of the hazard analysis allows us to account for censored observations in the sample. The censoring arises from the fact that at each transition there are women who have 'not exited' — for these women, the observed duration is the entire time period between the birth of child i and the survey date. The reason for using a hazard model is that a hazard model enables us to account for this censoring, which would have been difficult using least squares techniques.

In this paper we characterise the baseline hazard function as a gamma distribution, which is the most general parameterisation of the baseline hazard function. The associated hazard function is very flexible and allows for a large number of possible shapes, including as special cases the Weibull, the exponential and the log normal distributions. This flexibility is a useful feature for this study as the hazard of having an additional child could increase to begin with and then decrease. The hazard rate is therefore non-monotonic and the gamma distribution has the flexibility to model this non-monotonic relationship.

We consider three different transitions. Transition 1 → 2 denotes the duration between child 1 and child 2, transition 2 → 3 denotes the duration between child 2 and child 3 and, finally, transition 3 → 4 the duration between child 3 and child 4. We first study the entire sample and then consider separate regressions for each of the four races. Transitions beyond 3 → 4 are not examined because the sample size drops significantly.

The estimated equation is therefore given by

$$\ln(DURAT_i) = \beta X + \varepsilon. \tag{1}$$

Here, $DURAT_i$ is the duration between child i and i + 1, 2, 3. However, if any woman stops at i, then the observed duration is the entire period from the year in which she had her ith child and the survey year. These are the women who are 'censored'. Here, X denotes the set of explanatory variables and ε the set of unobservables that affect duration between successive births.

The explanatory variables used in the regression for each transition (X) include a set of sex composition variables and a set of other control variables that are likely to affect son preference. The advantage of including these other control variables (other than the sex-composition variables) in the set of explanatory variables is that they allow us to examine the effect of parental/household characteristics that could potentially affect son preference. The observed effects of these variables could have important policy implications. The set of sex-composition variables include: the total number of existing sons (TOTMAL); a dummy to indicate whether the existing children are of the same sex or not (DIFFSEX = 1 if the existing children are of different sex, 0 otherwise); and a dummy for the sex of the previous child (SEXPREV = 1 if the previous child is male, 0 otherwise). The other control variables that are included in the set of explanatory variables are: the age of the mother at previous birth (AGEPREV);[8] the age of the mother at the time of the survey (AGEMOTH); a dummy to indicate

[8] In the case of Transition 1 → 2, AGEPREV refers to the age of the mother at the time of birth of the first child; in the case of Transition 2 → 3, AGEPREV refers to the age of the mother at the time of birth of the second child; and in the case of Transition 3 → 4, AGEPREV refers to the age of the mother at the time of birth of the third child.

whether the woman lives in a rural area or not (RURAL = 1 if the household lives in a rural area, 0 otherwise); and three dummy variables to indicate the highest level of education attained by the mother (EDUCM1, EDUCM2, EDUCM3).[9] SEXPREV, TOTMAL and DIFFSEX are the 'gender preference variables'. The mother's education dummies are measures of the parental quality endowment. We examine the robustness of the results by re-estimating equation (1) using the highest level of education attained by the household head and current household income as measures of parental (household) endowment. Possible racial differences are examined by the inclusion of a set of race dummies (to account for differences in the intercept) and a set of interaction terms where we interact the other explanatory variables with the race dummies (to account for differences in the slopes). The race dummies that we include are BLACK, COLOURED and WHITE. The reference category is that the household is Indian. Finally, we include a set of province dummies to account for any other unobserved heterogeneity. The reference category is that the household lives in Transvaal.[10] See Table 1 for a description of all the explanatory variables used.

One of the problems with this empirical specification is that it could introduce potential collinearity and, hence, lead to inefficient estimates, since the estimating equation (1) could be over-specified.[11] We therefore consider alternative specifications. We start with the most parsimonious specification, where none of the gender preference variables (SEXPREV, TOTMAL and DIFFSEX) are included as explanatory variables. In this specification, the explanatory variables that are included are the three race dummies (BLACK, COLOURED, WHITE), the age of the mother at the time of the survey (AGEMOTH), the age of the mother at the time of the previous birth (AGEPREV), dummy for rural residence (RURAL) and the province dummies. This is Specification 1. We then include the three gender preference variables one at a time (Specifications 2–4) and then all of them

[9] EDUCM1 = 1 if the highest education attained by the mother is completed primary school (but less than secondary school). EDUCM2 = 1 if the highest level of schooling attained by the mother is more than primary school but less than secondary school. EDUCM3 = 1 if the highest level of schooling attained by the mother is more than secondary school. The reference category is that the mother has no schooling.

[10] Note that the province dummies used in the analysis correspond to the pre-1994 provinces. They have since changed.

[11] We would like to thank an anonymous referee for pointing this out to us.

Table 1: *Description of Variables Used*

Variable	Description
Non-interaction variables only	
DURAT$_i$	Duration between child *i* and child *i* + 1; *i* = 1, 2, 3
BLACK	=1 if the household is Black, 0 otherwise
COLOURED	=1 if the household is Coloured, 0 otherwise
INDIAN[a]	=1 if the household is Indian, 0 otherwise
WHITE	=1 if the household is White, 0 otherwise
AGEMOTH	Current age of mother
AGEPREV	Age of mother at the time of previous birth
SEXPREV	=1 if the previous child is a boy, 0 otherwise
TOTMAL	Total number of existing sons
DIFFSEX	=1 if the previous children are of different sex, 0 otherwise
SAMESEX	=1 if the previous children are of same sex, 0 otherwise
ATLEAST1B	=1 if the previous children contains at least one boy, 0 otherwise
RURAL	=1 if the mother lives in a rural area, 0 otherwise
INC1	=1 if the household income is in the bottom third of the income distribution, 0 otherwise
INC2	=1 if the household income is in the middle third of the income distribution, 0 otherwise
INC3[a]	=1 if the household income is in the top third of the income distribution, 0 otherwise
EDUCM0[a]	=1 if the mother has no education, 0 otherwise
EDUCM1	=1 if the highest level of education attained by the mother is some primary school, 0 otherwise
EDUCM2	=1 if the highest level of education attained by the mother is completed primary school, 0 otherwise
EDUCM3	=1 if the highest level of education attained by the mother is completed secondary school, 0 otherwise
EDUCHD0[a]	=1 if the household head has no education, 0 otherwise
EDUCHD1	=1 if the highest level of education attained by the household head is some primary school, 0 otherwise
EDUCHD2	=1 if the highest level of education attained by the household head is completed primary school, 0 otherwise
EDUCHD3	=1 if the highest level of education attained by the household head is completed secondary school, 0 otherwise

[a]Reference dummy.

together (Specification 5). Finally, we include the dummies for the highest level of schooling attained by the mother, EDUCM1, EDUCM2 and EDUCM3 (Specification 6). Next, we repeat the same sequence but include interaction of the race dummies with all the variables (except the province dummies). This gives us six more specifications. Specifiction 12 is therefore the most complete specification. We present the set of estimates for Specification 12 in Table 5 and the full set of estimates (for Specifications 1–12) is presented in the Appendix in Tables A1–A3.[12]

To examine the son preference hypothesis we need to look at the coefficients of SEXPREV and TOTMAL. If there is indeed a preference for sons over daughters, we expect to see the following:

1. at every transition we should observe an increased duration between children following the birth of a son, i.e., the coefficient of SEXPREV is positive;
2. the higher the number of existing boys, the greater is the duration between each successive child, which in turn implies that the coefficient of TOTMAL is positive.

3. Data and Descriptive Statistics

The data set used in this paper is from the 1993 South African Integrated Household Survey (SIHS), which is a part of the World Bank's Living Standard Measurement Study (LSMS) in a number of developing countries. In South Africa, the survey was conducted in 1993 jointly by the World Bank and the South Africa Labour and Development Research Unit (SALDRU) at the University of Cape Town. This cross-sectional data set is unique because it is the first survey that covers the entire South African population, including those in the predominantly Black 'homelands'.[13] The sample consists

[12] In order to examine whether there are any fertility effects, we re-estimated equation (1) by including the number of pregnancies (the best measure of fertility that is available) using a set of number-of-pregnancy dummies for each woman interacted with the race dummies. However, neither the pregnancy dummies nor the interaction dummies were statistically significant. Moreover, the associated standard errors were very high. We do not present these results for two reasons. First, the number of pregnancies is actually a measure of fecundity as opposed to fertility. Secondly, the number of pregnancies could be endogenous and could be correlated with the unobserved determinants of duration between births. It could also have effects on the bias of the other regressors. We would like to thank an anonymous referee for this point.

[13] The 'homelands' were designated residential regions for the Blacks during the apartheid regime. These were autonomous states within South Africa.

of approximately 9,000 households drawn randomly from 360 clusters. The questionnaire and summary statistics are contained in SALDRU (1994).

The questionnaire did not ask women directly about their fertility history and therefore the detailed history on child bearing had to be constructed from available data. In the survey, every member of the household was asked the identification code of his/her mother. Using this information we matched each mother with all her children. Hence, we were able to obtain the childbearing history of women who had at least one child who was alive at the time of the survey. However, because of the way the data were constructed, we were able to obtain the child characteristics for only the children who were alive at the time of the survey. In particular, we have no information on the children who had died or were living away from home and, hence, the duration between successive births ($DURAT_i$) could be measured with error. If the measurement error were truly random, then we could proceed with the estimation ignoring the measurement error, but keeping in mind that the intercept would need to be re-interpreted and the standard errors would need to be computed robustly to account for arbitrary heteroskedasticity. This is the approach we use in this paper. Let $DURAT_i$ be the observed duration and let $DURAT^*_i$ denote the true duration. The relationship between the observed and true duration can be written as:

$$DURAT_i = DURAT^*_i * v_i.$$

In the presence of measurement error, observed duration is always greater than the true duration. Now v_i is distributed over the range $[1,v]$, where v is the maximum potential duration between two successive births. So

$$\ln(DURAT_i) = \ln(DURAT^*_i) + \ln(v_i)$$

and comparing with equation (1) the only difference is that the error term is now $\varepsilon + \ln(v_i)$. Problems arise, however, if the measurement error is not random and is correlated with some of the regressors.[14] For example, child fostering or child mortality (both of which lead to increased observed duration) could be correlated with the exogenous

[14] We would like to thank an anonymous referee for drawing our attention to this issue.

variables (such as the sex of the child). In this case, the measurement error would be systematic and failure to account for this error could produce biased estimates of the exogenous variables.[15] Econometrically, we can do very little given the data at our disposal. However, for the Indian households measurement error of this kind is not likely to be a big problem. The average number of child deaths for Indian women in the sample is only 0.06. Previous studies have revealed that crude mortality rates, neo-natal mortality rates and infant mortality rates were lower for Indians than for Africans or Coloureds with the same basic environmental handicaps of poor housing and low income (see Kuper, 1960).[16] In addition, in Indian households (unlike in Black households) child fostering is not very common. Therefore, for the Indian women the results are unlikely to be biased. Finally, note that the observed sex ratios at birth (presented in Table 2) are not very different from the biological sex ratio of 105 (see Johansson and Nygren, 1991). While the (potentially) non-random measurement error could be an important problem, in this paper we do not attempt to correct for this bias. One must therefore be careful in interpreting the coefficients.

Table 2: *Selected Descriptive Statistics*

	All Households	Black	Coloured	Indian	White
Sample size (Transition 1 → 2)	8839	6981	756	262	840
Sample size[a] (Transition 2 → 3)	5763	4572	494	187	510
Sample size[a] (Transition 3 → 4)	3528	2971	294	93	170
Number of pregnancies	3.21	3.34	3.04	2.87	2.55
Average income (R)	2011.64	963.21	1767.94	3347.73	25250.93
Duration between child 1 and child 2	41.715 (33.996)	43.048 (35.902)	39.221 (27.292)	36.981 (27.865)	34.358 (21.382)

[15] Rose (1999), in her study of infant mortality in rural India, finds that there is a significant and systematic under-reporting of female births and failure to account for this reporting bias results in biased estimates.

[16] Yach (1988) reports infant mortality rates for the different races in South Africa over the period 1981–5 as follows: between 94 and 124 per 1000 live births for the Blacks; 51.9 per 1000 live births for the Coloureds; 17.9 per 1000 live births for Indians; and 12.3 per 1000 live births for Whites.

Table 2: *Continued*

	All Households	Black	Coloured	Indian	White
Duration between child 2 and child 3[a]	46.279	46.332	46.531	47.613	44.188
	(33.136)	(33.947)	(29.390)	(28.000)	(27.083)
Duration between child 3 and child 4[a]	44.234	44.024	47.652	41.189	45.081
	(29.828)	(29.815)	(30.618)	(28.077)	(29.449)
First child boy	0.509	0.508	0.485	0.511	0.538
	(0.500)	(0.500)	(0.500)	(0.501)	(0.499)
Second child boy[a]	0.506	0.510	0.486	0.551	0.478
	(0.500)	(0.500)	(0.500)	(0.499)	(0.500)
Third child boy[a]	0.520	0.516	0.554	0.527	0.518
	(0.500)	(0.500)	(0.498)	(0.502)	(0.501)
Total number of existing sons at Transition $2 \rightarrow 3$[a]	1.019	1.020	0.994	1.048	1.022
	(0.711)	(0.708)	(0.711)	(0.743)	(0.729)
Total number of existing sons at Transition $3 \rightarrow 4$[a]	1.545	1.546	1.571	1.376	1.559
	(0.894)	(0.890)	(0.913)	(0.920)	(0.923)
Existing children of different sex at Transition $2 \rightarrow 3$[a]	0.670	0.672	0.671	0.607	0.677
	(0.470)	(0.470)	(0.470)	(0.489)	(0.468)
Existing children of different sex at Transition $3 \rightarrow 4$[a]	0.890	0.884	0.886	0.893	0.939
	(0.313)	(0.320)	(0.318)	(0.310)	(0.239)
Age of mother at the time of birth of first child	22.993	22.699	23.148	23.527	25.130
	(6.548)	(6.714)	(6.174)	(5.027)	(5.396)
Age of mother at the time of birth of second child[a]	26.802	26.858	26.350	26.059	27.012
	(6.784)	(7.145)	(5.866)	(4.631)	(4.560)
Age of mother at the time of birth of third child[a]	29.896	30.039	29.133	29.043	29.182
	(6.899)	(7.173)	(5.688)	(5.077)	(4.099)
Highest level of education attained by the mother is some primary school	0.327	0.362	0.358	0.183	0.052
	(0.469)	(0.481)	(0.480)	(0.388)	(0.223)
Highest level of education attained by the mother is completed primary school	0.283	0.262	0.433	0.393	0.286
	(0.450)	(0.440)	(0.496)	(0.489)	(0.452)
Highest level of education attained by the mother is completed secondary school	0.186	0.141	0.134	0.317	0.573
	(0.389)	(0.348)	(0.340)	(0.466)	(0.495)
Rural residence	0.565	0.698	0.062	0.004	0.086
	(0.496)	(0.459)	(0.242)	(0.062)	(0.280)
Sex ratio at birth	104.186	104.201	99.483	111.719	106.369

[a]Computed for the non-censored sample.
Figures in parentheses are standard deviations.

The childbearing history was obtained for a subset of the women surveyed. The sample used in this paper consisted of 8,839 women, of whom 6,981 (78.97%) were Black, 756 (8.55%) were Coloured, 262 (2.96%) were Indian and 840 (9.51%) were White. In terms of sample proportions, this sample is fairly representative of the population distribution of South Africa. The average number of pregnancies was 3.34 per Black woman, 3.04 per Coloured woman, 2.87 per Indian woman and 2.54 per White woman.[17] The Blacks were the poorest (mean household income R963.21), followed by the Coloured (mean household income R1767.94), Indian (mean household income R3347.73) and White (mean household income R25250.93). We also compute the sex ratios at birth for the sample that we use in this paper. The ratio for the full sample is 104.186, which is actually close to the biological ratio of approximately 105. While there is some evidence of racial difference in the sex ratios — they range from 99.483 for the Coloured households to 111.719 for the Indian households — reported sex ratios at birth do not, however, display any unusual male bias. See Table 2 for descriptive statistics for selected variables.

Table 3 presents the average duration between births at each transition, conditional on the sex of the previous children. Notice that for Indian households, the average duration between child 1 and child 2 is significantly higher if the first child is a boy — 47.23 months compared to 35.74 months if the first child is a girl. The average duration is not significantly different controlling for the sex of the first child for women belonging to the other races. The average duration between child 2 and child 3 is monotonically increasing in the number of existing sons for the Indian women — the average duration between child 2 and child 3 is 41.68 months when both child 1 and child 2 are girls, is 49.54 months when one of child 1 and child 2 is a boy and is 54.5 months when both child 1 and child 2 are boys. This monotonic relationship between the sex of the existing children and the duration between children does not, however, hold for non-Indian women. For Coloured and White women, the average duration is the highest if child 1 and child 2 are of different sex, while for Black women the average duration is the lowest if child 1 and child 2 are of different sex.

The parity progression ratios presented in Table 4 show that 65.5%

[17] Under the assumption that no pregnancy is unwanted, the number of pregnancies per woman is, in our opinion, the best available measure of fertility.

Table 3: *Average Duration between Children by Race and Sex of Children*

No. of children	Sex of existing children	Black	Coloured	Indian	White
1	1 daughter	53.81	46.57	35.74	38.52
	1 son	53.15	47.47	47.23	39.34
	t-value for difference in duration	0.55	–0.33	–2.51*	–0.35
2	2 daughters	48.37	43.38	41.68	44.4
	1 son, 1 daughter	44.64	51.79	49.54	48.76
	2 sons	47.62	41.06	54.5	38.95
	children of different sex	44.64	51.79	49.55	48.76
	children of same sex	47.98	42.19	46.65	41.5
	t-value for difference in duration[a]	–2.68*	2.82*	0.64	1.70**
	no boy	48.37	43.38	41.68	44.4
	at least one boy	45.69	47.67	51.71	44.1
	t-value for difference in duration[b]	1.84**	–1.1	–1.72**	0.07
3	3 daughters	45.95	49.5	33.6	30
	1 son, 2 daughters	44.48	54.49	44	59
	2 sons, 1 daughter	43.73	47	39.27	43
	3 sons	42.13	35.43	44	37.71
	children of different sex	44.11	49.86	42	51
	children of same sex	43.79	42.44	39.27	34.15
	t-value for difference in duration[a]	0.2	–1.3	0.37	1.71**
	no boy	45.95	49.8	33.6	30
	at least one boy	43.77	47.29	42.38	48
	t-value for difference in duration[b]	1.02	0.34	–0.65	–1.39

[a]Difference in mean for SAMESEX = 1 and SAMESEX = 0. See definition of SAMESEX dummy in Table 1.
[b]Difference in mean for ATLEAST1B = 1 and ATLEAST1B = 0. See definition of ATLEAST1B dummy in Table 1.
*Significant at the 95% level; **significant at the 90% level.

Black, 65.3% Coloured, 71.4% Indian and 60.6% White women go on to have a second child.[18] However, the proportion of women who have a third child drops significantly for the Coloured, Indian and White women — of the women who had more than one child, only 59.5% of Coloured 49.7% of Indian and 33.3% of White women went on to have

[18] Parity progression refers to the percentage who go on to the next level — in this case, the percentage who have an additional child at each transition.

Table 4: *Proportion of Women who have another Child by Race and Sex of Children*

No. of children	Sex of existing children	Black	Coloured	Indian	White
1	parity progression[c]	65.5	65.3	71.4	60.6
	1 daughter	65.2	63.4	74	60.6
	1 son	65.8	67.8	70.5	62.2
	t-value for difference in proportion	–0.472	–1.71**	0.722	–0.41
2	parity progression[c]	65	59.5	49.7	33.3
	2 daughters	65.1	61.9	80.9	38.3
	1 son, 1 daughter	64.3	55.2	38.1	26.6
	2 sons	67.9	67.2	42.9	39.6
	children of different sex	64.2	54.3	36.9	26.4
	children of same sex	65.8	64.7	60.2	39.5
	t-value for difference in proportion[a]	–1.132	–2.348*	–3.168*	–3.137*
	no boy	64.51	61.9	80.85	38.46
	at least one boy	65.13	58.7	39.29	31.58
	t-value for difference in proportion[b]	–0.37	0.63	4.93*	1.44
3	parity progression[c]	62.5	46.9	39.8	21.8
	3 daughters	59.8	52.6	31.3	25
	1 son, 2 daughters	65.5	37.8	39.5	22.6
	2 sons, 1 daughter	62.5	56	42.9	18.5
	3 sons	65.7	45.7	50	28
	children of different sex	62.7	46	40	20.2
	children of same sex	61.8	47.7	39.3	25.5
	t-value for difference in proportion[a]	0.49	–0.26	0.06	–0.77
	no boy	59.29	52.63	31.25	25
	at least one boy	62.92	46.09	41.56	21.23
	t-value for difference in proportion[b]	–1.34	0.75	–0.77	0.41

[a]Difference in proportion for SAMESEX = 1 and SAMESEX = 0. See definition of SAMESEX dummy in Table 1.
[b]Difference in proportion for ATLEAST1B = 1 and ATLEAST1B = 0. See definition of ATLEAST1B dummy in Table 1.
[c]Parity progression refers to the percentage who go on to the next level — in this case, the percentage who have an additional child at each transition.
*Significant at the 95% level; **significant at the 90% level.

a third child. We find an even larger drop in the proportion of women who had a fourth child. Table 4 also presents parity progression ratios conditional on the sex-mix of existing children. Except for Coloured women, the parity progression ratio is not different conditional on the sex of the first child. For Coloured women, the proportion of women who had a second child is significantly higher if the first child was a boy. The proportion of women who had a third child is the highest among the Blacks and the lowest among the Whites. The proportion of Coloured, Indian and White women who had a third child is significantly higher if the first two children are of the same sex. Notice also that the proportion of Indian women who had a third child is significantly higher if both child 1 and child 2 were girls. In the case of Transition 3 → 4, the sex-mix of the existing children does not significantly affect the parity progression ratios.[19]

4. Results

Table 5 presents the accelerated hazard regression results corrected for arbitrary heteroskedasticity for the joint estimation (for all races) at each transition. The second column presents the transition from the first to the second child (1 → 2); the third column presents the transition from the second to the third child (2 → 3); and the fourth column presents the transition from the third to the fourth child (3 → 4). A negative sign on the coefficient decreases the duration between successive births (and increases the hazard of having a subsequent child), while a positive sign increases the duration between successive births (and decreases the hazard of having a subsequent child). Note that at each transition the estimated acceleration factor is given by e^{β}. The acceleration factor helps in isolating the magnitude of the effect of a particular variable on duration. If the acceleration factor is greater than unity, then that variable increases the duration and if it is less than unity, then it decreases the duration between children.[20]

[19] The problem with using parity progression ratios as a measure of the desire to have another child is that it is very difficult to throw much light on desired fertility from observed behaviour. In particular, the major problem arises in the analysis of 'complete versus incomplete families'. To account for this problem, we can either examine the parity progressions for women beyond a certain age — for example Haughton and Haughton (1998) constrain their sample to include women who are more than 37.4 years old — or use a cohort level analysis. Either way, the sample size becomes very small in some cases. In this paper, therefore, we focus only on the duration between successive births.

[20] The full set of estimates for the alternative specifications (Specifications 1–12) is presented in the Appendix (Tables A1–A3). Notice that the results on son preference (the sign and significance of SEXPREV and TOTMAL) actually improve when we include the interaction terms and control for the educational attainment of the mother (Specification 6, for Transition 1 → 2 and Specification 12 for Transitions 2 → 3 and 3 → 4).

Table 5: *Accelerated Hazard Regressions for Duration between Successive Children*

	Transition 1 → 2 (distribution: gamma)		Transition 2 → 3 (distribution: gamma)		Transition 3 → 4 (distribution: gamma)	
	Coef.	Accel'n factor	Coef.	Accel'n factor	Coef.	Accel'n factor
SEXPREV	0.048*	1.050	0.107**	1.113	0.037	1.037
	(0.021)		(0.056)		(0.054)	
SEXPREV*BLACK	–0.055*	0.946	–0.103**	0.902	–0.039	0.962
	(0.022)		(0.057)		(0.054)	
SEXPREV*COLOURED	–0.047**	0.954	–0.037	0.964	–0.041	0.960
	(0.025)		(0.062)		(0.062)	
SEXPREV*WHITE	–0.049*	0.952	–0.093	0.911	–0.018	0.982
	(0.025)		(0.066)		(0.074)	
TOTMAL			0.097*	1.102	–0.014	0.986
			(0.034)		(0.032)	
TOTMAL*BLACK			–0.097*	0.907	0.009	1.009
			(0.034)		(0.032)	
TOTMAL*COLOURED			–0.072*	0.931	–0.012	0.988
			(0.038)		(0.036)	
TOTMAL*WHITE			–0.101*	0.904	0.021	1.021
			(0.041)		(0.041)	
DIFFSEX			0.059**	1.061	–0.008	0.992
			(0.034)		(0.053)	
DIFFSEX*BLACK			–0.071*	0.932	0.005	1.005
			(0.035)		(0.054)	
DIFFSEX*COLOURED			–0.009	0.991	0.038	1.038
			(0.038)		(0.059)	
DIFFSEX*WHITE			–0.002	0.998	0.100	1.106
			(0.041)		(0.070)	
BLACK	0.110	1.116	0.011	1.011	0.386*	1.471
	(0.080)		(0.120)		(0.183)	
COLOURED	0.091	1.096	0.002	1.002	0.429*	1.536
	(0.090)		(0.135)		(0.205)	
WHITE	–0.104	0.901	–0.094	0.911	0.066	1.068
	(0.088)		(0.145)		(0.283)	
AGEPREV	–0.003	0.997	0.003	1.003	0.016*	1.016
	(0.003)		(0.005)		(0.007)	
AGEPREV*BLACK	0.000	1.000	–0.003	0.997	–0.014*	0.987
	(0.003)		(0.005)		(0.007)	
AGEPREV*COLOURED	0.002	1.002	0.000	1.000	–0.009	0.991
	(0.003)		(0.005)		(0.007)	
AGEPREV*WHITE	0.004	1.004	0.007	1.007	–0.001	0.999
	(0.003)		(0.005)		(0.010)	

Table 5: *Continued*

	Transition 1 → 2 (distribution: gamma)		Transition 2 → 3 (distribution: gamma)		Transition 3 → 4 (distribution: gamma)	
	Coef.	Accel'n factor	Coef.	Accel'n factor	Coef.	Accel'n factor
AGEMOTH	0.004*	1.004	–0.001	0.999	0.000	1.000
	(0.001)		(0.002)		(0.004)	
AGEMOTH*BLACK	0.000	1.000	0.003	1.003	0.000	1.000
	(0.001)		(0.002)		(0.004)	
AGEMOTH*COLOURED	–0.003**	0.997	0.001	1.001	–0.002	0.998
	(0.002)		(0.003)		(0.004)	
AGEMOTH*WHITE	0.000	1.000	0.000	1.000	–0.001	0.999
	(0.002)		(0.003)		(0.006)	
RURAL	1.111	3.039	–0.049	0.952	–0.078	0.925
	(36.485)		(0.036)		(0.064)	
RURAL*BLACK	–1.125	0.325	0.026	1.026	0.054	1.055
	(36.485		(0.037)		(0.065)	
RURAL*COLOURED	–1.149	0.317				
	(36.485)					
RURAL*WHITE	–1.121	0.326	0.006	1.006	0.078	1.082
	(36.485)		(0.052)		(0.084)	
EDUCM1	0.023	1.023	0.062	1.064	–0.034	0.967
	(0.044)		(0.065)		(0.097)	
EDUCM1*BLACK	–0.036	0.964	–0.063	0.939	0.037	1.038
	(0.044)		(0.065)		(0.097)	
EDUCM1*COLOURED	–0.003	0.997	–0.091	0.913	0.017	1.017
	(0.052)		(0.074)		(0.109)	
EDUCM1*WHITE	–0.007	0.993	–0.163**	0.849	0.061	1.063
	(0.056)		(0.087)		(0.154)	
EDUCM2	–0.006	0.994	0.013	1.013	0.052	1.054
	(0.042)		(0.063)		(0.101)	
EDUCM2*BLACK	0.005	1.005	0.010	1.010	–0.032	0.968
	(0.042)		(0.064)		(0.102)	
EDUCM2*COLOURED	–0.021	0.979	–0.021	0.980	–0.045	0.956
	(0.050)		(0.073)		(0.114)	
EDUCM2*WHITE	0.053	1.054	–0.044	0.957	–0.088	0.916
	(0.049)		(0.078)		(0.141)	
EDUCM3	–0.003	0.997	0.029	1.029	0.053	1.054
	(0.045)		(0.065)		(0.108)	
EDUCM3*BLACK	0.028	1.029	0.013	1.013	–0.017	0.983
	(0.045)		(0.066)		(0.109)	
EDUCM3*COLOURED	0.007	1.007	–0.063	0.939	–0.118	0.889
	(0.056)		(0.079)		(0.125)	

Table 5: *Continued*

	Transition 1 → 2 (distribution: gamma)		Transition 2 → 3 (distribution: gamma)		Transition 3 → 4 (distribution: gamma)	
	Coef.	Accel'n factor	Coef.	Accel'n factor	Coef.	Accel'n factor
EDUCM3*WHITE	0.021	1.021	–0.039	0.962	–0.025	0.976
	(0.050)		(0.078)		(0.145)	
CONSTANT	1.198*		1.277*		0.895*	
	(0.079)		(0.119)		(0.182)	
κ	0.616*		0.618*		0.657*	
	(0.032)		(0.040)		(0.053)	
σ	0.149		0.153		0.145	
	(0.002)		(0.002)		(0.003)	
Number of observations	8839		5763		3528	
Number censored	3076		2235		1460	
Log likelihood	1275	0.3901	388	0.3287	175	0.0363

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Coef., coefficient; Accel'n, acceleration.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution.

Let us first consider Transition 1 → 2. In this case, the only gender preference variable that we include in our regression is SEXPREV. For son preference to exist, the coefficient of SEXPREV must be positive. The regression results presented show that not only is the SEXPREV dummy positive and statistically significant, the interaction terms (SEXPREV*BLACK, SEXPREV*COLOURED and SEXPREV*WHITE) are all negative and statistically significant. The configuration of signs implies that, controlling for other characteristics, the duration between child 1 and child 2 is higher if the first child is a male child and this duration is significantly higher for Indian households compared to households belonging to other races. Note that none of the three race dummies are statistically significant implying that, *per se*, there is no racial difference in the duration following the birth of the first child — the racial difference comes into effect when we control for the sex of the first child. Turning to the other results, AGEMOTH is positive and statistically significant, which implies that the higher the age of the mother at the time of the survey, the higher is the duration between child 1 and child 2. This result is somewhat surprising. One reason could be measurement error in the method of constructing the duration between successive children, as older mothers are more likely to have children living away from home. Of course there could be other reasons as well — for example, older women might find it more difficult to become pregnant or the opportunity costs of taking time off work to have a child are greater for older women.

Let us now turn to the regression results for Transition 2 → 3. In this case we include all three gender preference variables (SEXPREV, TOTMAL and DIFFSEX) and we also include the interaction of each of these variables with the race dummies. SEXPREV is positive and statistically significant and the three interaction terms (SEXPREV*BLACK, SEXPREV*COLOURED and SEXPREV*WHITE) are all negative, though they are not always statistically significant. This configuration of signs implies that the duration between child 2 and child 3 is higher for Indian households if child 2 is a boy. The sign configurations for TOTMAL, TOTMAL*BLACK, TOTMAL*COLOURED and TOTMAL*WHITE are exactly the same — TOTMAL is positive and statistically significant, while all of the interaction terms are negative and statistically significant. This implies that an increase in the number of sons a woman has significantly reduces the hazard and increases the duration between child 2 and child 3 for Indian women. Finally, DIFFSEX is positive and statistically significant and the three interaction terms (DIFFSEX*BLACK, DIFFSEX*COLOURED and DIFFSEX*WHITE) are negative, though not always statistically significant.

Turning to Transition 3 → 4, we see that none of the gender preference variables are statistically significant, though the signs are similar to those in Transitions 1 → 2 and 2 → 3. However, note that the coefficient estimates of DIFFSEX and the interaction terms (DIFFSEX*BLACK, DIFFSEX*COLOURED and DIFFSEX*WHITE) all change signs as we move from Transition 2 → 3 to Transition 3 → 4, though none of the coefficients are statistically significant. Interestingly, in this case two of the race dummies (BLACK and COLOURED) are positive and statistically significant. This implies that, relative to Indian households, the duration between the third and the fourth child is significantly higher for Black and Coloured households.

Surprisingly, once again we find that the higher the age of the mother at the time of the previous birth, the greater the duration.

The duration analysis for the entire sample suggests that there are significant racial differences in son preference. In particular it is clear that the Indian households behave quite differently from the others. When we jointly estimate the duration across all races, we are in effect imposing a restriction that the scale and shape parameters for the distribution (κ,σ) are the same for all races, i.e., $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4$ and $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$, where the subscripts denote the race of the household (1 = Black, 2 = Coloured, 3 = Indian and 4 = White). However, the null hypothesis is rejected using a likelihood ratio test. Hence, it is more efficient to estimate the duration between successive children separately by race. These results are presented in Tables 6–9 for Black, Coloured, Indian and White households, respectively. While the results are similar to those presented in Table 5, separating the analysis by race further strengthens the argument that son preference is significant for Indian households. For the Indian households (Table 8), the coefficient estimate of SEXPREV is positive and statistically significant for Transition 1 $\rightarrow$ 2 and for Transition 2 $\rightarrow$ 3 and is positive, though not statistically significant, for Transition 3 $\rightarrow$ 4. Therefore, irrespective of the number of children, Indian women delay having another child following the birth of a son. The coefficient estimate of TOTMAL is positive and statistically significant for Transition 2 $\rightarrow$ 3 and is positive, though not statistically significant, for Transition 3 $\rightarrow$ 4. Therefore, the greater the number of existing sons, the greater is the duration between successive children. Finally, the coefficient estimate of DIFFSEX is always negative and is negative, but not statistically significant, for Transition 3 $\rightarrow$ 4. Indian women therefore choose to reduce the duration between children if the existing children are of different sexes. The predicted median duration between child 1 and child 2 is 48.5 months if the first child is a boy and 45.8 months if the first child is a girl; the median duration between child 2 and child 3 is 48.1 months if the second child is a boy and 44.8 months if the second child is a girl and, finally, the median duration between child 3 and child 4 is 50.4 months if the third child is a boy and is 49.5 months if the third child is a girl. Likewise, the predicted median duration between child 2 and child 3 is 43.8 months if both the first and the second children are girls, is 46.8 months if only one of the first two children is a girl and is 47.8 months if both child 1 and child 2 are boys. The predicted durations between child 3 and child 4 are 49.8, 49.6, 50.0

Table 6: *Accelerated Hazard Regressions for Duration between Successive Children — Black Households Only*

	Transition 1 → 2 (distribution: gamma)		Transition 2 → 3 (distribution: gamma)		Transition 3 → 4 (distribution: gamma)	
	Coef.	Accel'n factor	Coef.	Accel'n factor	Coef.	Accel'n factor
SEXPREV	–0.007	0.993	–0.004	0.996	–0.002	0.998
	(0.004)		(0.008)		(0.008)	
TOTMAL			0.000	1.000	–0.005	0.995
			(0.006)		(0.005)	
DIFFSEX			–0.012*	0.988	–0.003	0.997
			(0.006)		(0.007)	
AGEPREV	–0.003*	0.997	0.001	1.001	0.003*	1.003
	(0.000)		(0.001)		(0.001)	
AGEMOTH	0.004*	1.004	0.002*	1.002	0.001*	1.001
	(0.000)		(0.000)		(0.000)	
RURAL	–0.014*	0.986	–0.025*	0.975	–0.025*	0.975
	(0.007)		(0.009)		(0.010)	
EDUCM1	–0.014*	0.986	0.000	1.000	0.003	1.003
	(0.006)		(0.007)		(0.008)	
EDUCM2	–0.002	0.998	0.024*	1.024	0.020*	1.021
	(0.007)		(0.009)		(0.010)	
EDUCM3	0.025*	1.025	0.042*	1.043	0.036*	1.037
	(0.009)		(0.012)		(0.016)	
CONSTANT	1.308*		1.283*		1.281*	
	(0.013)		(0.017)		(0.022)	
κ	0.613*		0.564*		0.656*	
	(0.036)		(0.045)		(0.056)	
σ	0.150		0.156		0.148	
	(0.002)		(0.003)		(0.003)	
Number of observations	6981		4572		2971	
Number censored	2409		1601		1115	
Log likelihood	1038.3759		365.1923		178.5166	

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Coef., coefficient; Accel'n, acceleration.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution.

Table 7: *Accelerated Hazard Regressions for Duration between Successive Children — Coloured Households Only*

	Transition 1 → 2 (distribution: gamma)		Transition 2 → 3 (distribution: gamma)		Transition 3 → 4 (distribution: gamma)	
	Coef.	Accel'n factor	Coef.	Accel'n factor	Coef.	Accel'n factor
SEXPREV	0.001	1.001	–0.050*	0.951	–0.013	0.987
	(0.012)		(0.021)		(0.027)	
TOTMAL			0.023	1.023	–0.023	0.978
			(0.014)		(0.015)	
DIFFSEX			0.036*	1.037	0.027	1.028
			(0.014)		(0.023)	
AGEPREV	–0.001	0.999	0.000	1.000	0.006*	1.006
	(0.001)		(0.002)		(0.002)	
AGEMOTH	0.002*	1.002	0.002*	1.002	–0.001	0.999
	(0.001)		(0.001)		(0.002)	
RURAL	–0.037	0.964	–0.037	0.963	–0.007	0.993
	(0.027)		(0.030)		(0.047)	
EDUCM1	0.015	1.015	–0.030	0.971	–0.005	0.995
	(0.026)		(0.028)		(0.047)	
EDUCM2	–0.028	0.973	–0.001	0.999	0.018	1.018
	(0.027)		(0.030)		(0.049)	
EDUCM3	0.009	1.009	–0.034	0.967	–0.054	0.948
	(0.033)		(0.036)		(0.061)	
CONSTANT	1.289*		1.362*		1.276*	
	(0.048)		(0.059)		(0.093)	
κ	0.797*		1.493*		0.476*	
	(0.116)		(0.192)		(0.207)	
σ	0.140*		0.109*		0.133*	
	(0.006)		(0.009)		(0.010)	
Number of observations	756		494		294	
Number censored	262		200		156	
Log likelihood	113.6360		33.5077		20.9133	

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Coef., coefficient; Accel'n, acceleration.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution.

Table 8: *Accelerated Hazard Regressions for Duration between Successive Children — Indian Households Only*

	Transition 1 → 2 (distribution: Weibull)		Transition 2 → 3 (distribution: Weibull)		Transition 3 → 4 (distribution: Weibull)	
	Coef.	Accel'n factor	Coef.	Accel'n factor	Coef.	Accel'n factor
SEXPREV	0.050*	1.051	0.084**	1.088	0.061	1.063
	(0.021)		(0.047)		(0.049)	
TOTMAL			0.075*	1.078	0.024	1.024
			(0.029)		(0.029)	
DIFFSEX			–0.054**	1.055	–0.007	0.993
			(0.030)		(0.049)	
AGEPREV	–0.003	0.997	0.001	1.001	0.016*	1.016
	(0.003)		(0.004)		(0.006)	
AGEMOTH	0.004*	1.004	0.000	1.000	0.000	1.000
	(0.001)		(0.002)		(0.004)	
EDUCM1	0.038	1.039	0.069	1.071	–0.023	0.978
	(0.043)		(0.058)		(0.091)	
EDUCM2	0.000	1.000	0.036	1.036	0.062	1.064
	(0.041)		(0.056)		(0.095)	
EDUCM3	–0.003	0.997	0.050	1.052	0.040	1.041
	(0.044)		(0.057)		(0.101)	
CONSTANT	1.264*		1.305*		0.964*	
	(0.085)		(0.103)		(0.161)	
α	7.166*		8.513*		8.694*	
	(0.396)		(0.622)		(1.026)	
Number of observations	262		187		93	
Number censored	75		94		56	
Log likelihood	30.2307		20.9024		–5.3012	

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Coef., coefficient; Accel'n, acceleration.
*Significant at the 95% level; **significant at the 90% level.
α: Parameter of the Weibull distribution.

and 50.6 months for TOTMAL = 0, 1, 2 and 3, respectively. In each of these cases, the predicted median duration was computed for an Indian woman residing in Natal and having no schooling.

Table 9: *Accelerated Hazard Regressions for Duration between Successive Children — White Households Only*

	Transition 1 → 2 (distribution: Weibull)		Transition 2 → 3 (distribution: Weibull)		Transition 3 → 4 (distribution: Weibull)	
	Coef.	Accel'n factor	Coef.	Accel'n factor	Coef.	Accel'n factor
SEXPREV	–0.003	0.997	–0.011	0.989	0.023	1.023
	(0.012)		(0.029)		(0.041)	
TOTMAL			–0.010	0.990	–0.003	0.997
			(0.018)		(0.021)	
DIFFSEX			0.039*	1.040	0.102*	1.107
			(0.018)		(0.037)	
AGEPREV	0.002	1.002	0.003	1.003	0.009	1.009
	(0.002)		(0.003)		(0.006)	
AGEMOTH	0.004*	1.004	0.002	1.002	0.002	1.002
	(0.001)		(0.002)		(0.003)	
RURAL	–0.002	0.998	–0.059*	0.943	–0.067	0.935
	(0.021)		(0.030)		(0.047)	
EDUCM1	0.012	1.012	–0.095*	0.909	0.020	1.020
	(0.033)		(0.047)		(0.093)	
EDUCM2	0.042**	1.042	–0.029	0.971	–0.024	0.977
	(0.024)		(0.037)		(0.077)	
EDUCM3	0.013	1.013	–0.003	0.997	0.024	1.024
	(0.022)		(0.035)		(0.076)	
CONSTANT	1.077*		1.285*		1.010*	
	(0.041)		(0.065)		(0.177)	
κ	0.405*					
	(0.100)					
σ	0.146					
	(0.005)					
α			9.283*		10.874*	
			(0.479)		(1.218)	
Number of observations	840		510		170	
Number censored	330		340		133	
Log likelihood	103.5684		–3.3680		–7.8286	

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Coef., coefficient; Accel'n, acceleration.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution; α: Parameter of the Weibull distribution.

Interestingly, note that the coefficient of TOTMAL is never statistically significant for the Black, Coloured and White households and SEXPREV is statistically significant only for Transition 2 → 3 for Black households and in no other cases for the sample of non-Indian households. However, the coefficient of DIFFSEX is always significantly different from zero for Transition 2 → 3 and is positive and statistically significant for Transition 3 → 4 for White households. In the case of Transition 2 → 3 for White and Coloured households, the coefficient of DIFFSEX is positive and statistically significant, implying that the duration between child 2 and child 3 is higher if child 1 and child 2 are of different sexes. For White households, the coefficient of DIFFSEX is positive and statistically significant for Transition 3 → 4, implying that the duration between child 3 and child 4 is higher if the existing children are not of the same sex. For the Coloured households also, the coefficient estimate of DIFFSEX is positive, though not statistically significant in the case of Transition 3 → 4. For the Black households, the coefficient of DIFFSEX is always negative, though it is not statistically significant for Transition 3 → 4.

The highest level of education attained by the mother is used as a measure of the economic status of the household (measure of household endowment). The use of the educational attainment of the mother can be criticised on the basis that in most developing countries, women's educational attainment is not a true indicator of the economic status of the household. We therefore re-estimate the model using two alternative specifications. In the first we use the highest level of education attained by the household head as the measure of economic status of the household and in the second we use the current income level of the household as the relevant measure.[21] It must be noted, given the fact that the data are not retrospective, that both of these measures suffer from standard endogeneity problems and problems of measurement error. For example, current household income might incorrectly measure past household economic status and therefore incorrectly measure the effect of the household economic status on the duration between births. A similar problem arises with the educational attainment of the household head. In the absence of any variables that would be better approximations of the economic status of the

[21] It would be useful to control for fathers' educational attainment as well. However, marriage is not a social pre-condition for having children in this sample and more than 25% of the women in the sample who had children are not married (see Gangadharan and Maitra, 2001).

household at the time of the decision on birth spacing, we use household income and the educational attainment of the household head as explanatory variables.[22] Instead of using household income as a continuous variable, however, we include two dummies (INC1 and INC2) to determine the income category of the household.[23,24] We use three dummies to categorise the highest level of education attained by the household head — EDUCHD1, EDUCHD2 and EDUCHD3.[25] The results are presented in Tables 10 (all households) and 11 (Indian households).

The robustness results for all households presented in Table 10 are similar to those presented in Table 5. Irrespective of the number of existing children, the duration following the birth of a son is significantly higher for Indian households. The higher the number of existing sons, the greater is the duration between child 2 and child 3 for Indian households. Moreover, note that the results are similar irrespective of whether we use the highest level of education attained by the household head as our measure of household quality endowment or the income category dummies.

[22] Household assets, if available, could be used as a measure of household endowment. But again, the data on household assets that we have are at the time of the survey and not at the time of the birth of the child.

[23] INC1 = 1 if the household income is in the bottom third of the income distribution, 0 otherwise and INC2 = 1 if the household income is in the middle third of the income distribution, 0 otherwise. The reference category is that the household income is in the top third of the income distribution.

[24] As mentioned above, the use of household income as a measure of household endowment is problematic, particularly since the data are not retrospective in nature. Household income could therefore be endogenous. To account for this endogeneity we include income categories as the relevant explanatory variable of interest. However, as one of the referees correctly notes, this solution will work only if there is not much mobility across income groups. Maluccio *et al*. (2000), using a subset of this data set, show that 'there was substantial movement within the distribution over the period 1993-1998. Furthermore, no group was immune to change, although households in the lowest and highest quintiles were less likely to have changed category. More than one-half of the households in the lowest quintile in 1993 had moved up in the distribution by 1998. For the three middle quintiles, three-fourths of the households had transited to a different quintile.' In light of existing evidence, the assumption of stationary distribution might not be very good.

[25] EDUCHD1 = 1 if the highest level of schooling attained by the household head is some primary school. EDUCHD2 = 1 if the highest level of schooling attained by the household head is completed primary school (but less than secondary school). EDUCHD3 = 1 if the highest level of schooling attained by the household head is more than secondary school. The reference category is that the household head has no schooling.

Table 10: *Robustness — All Households*

	Transition 1 → 2		Transition 2 → 3		Transition 3 → 4	
	Educ'n	Income quantiles	Educ'n	Income quantiles	Educ'n	Income quantiles
SEXPREV	0.052*	0.050*	0.099**	0.099**	0.043	0.035
	(0.022)	(0.022)	(0.056)	(0.056)	(0.054)	(0.055)
SEXPREV*BLACK	–0.058*	–0.057*	–0.095**	–0.095**	–0.044	–0.038
	(0.022)	(0.022)	(0.056)	(0.056)	(0.055)	(0.055)
SEXPREV*COLOURED	–0.053*	–0.050*	–0.027	–0.030	–0.047	–0.036
	(0.025)	(0.025)	(0.061)	(0.061)	(0.062)	(0.062)
SEXPREV*WHITE	–0.051*	–0.050*	–0.100	–0.099	–0.024	–0.025
	(0.025)	(0.025)	(0.066)	(0.066)	(0.074)	(0.076)
TOTMAL			0.098*	0.098*	–0.021	–0.020
			(0.034)	(0.034)	(0.029)	(0.029)
TOTMAL*BLACK			–0.099*	–0.098*	0.015	0.015
			(0.034)	(0.034)	(0.030)	(0.029)
TOTMAL*COLOURED			–0.072**	–0.073**	–0.004	–0.006
			(0.038)	(0.038)	(0.034)	(0.033)
TOTMAL*WHITE			–0.111*	–0.109*	0.027	0.029
			(0.041)	(0.040)	(0.039)	(0.039)
DIFFSEX			0.051	0.050	0.010	0.001
			(0.034)	(0.034)	(0.051)	(0.050)
DIFFSEX*BLACK			–0.063	–0.062**	–0.013	–0.003
			(0.034)	(0.034)	(0.051)	(0.051)
DIFFSEX*COLOURED			–0.002	–0.001	0.019	0.021
			(0.038)	(0.038)	(0.057)	(0.057)
DIFFSEX*WHITE			0.007	0.004	0.083	0.094
			(0.041)	(0.041)	(0.069)	(0.068)
BLACK	0.131	0.122*	0.117	0.026	0.399*	0.297*
	(0.082)	(0.056)	(0.127)	(0.090)	(0.192)	(0.149)
COLOURED	0.082	0.089	0.094	–0.018	0.405**	0.320*
	(0.089)	(0.062)	(0.138)	(0.100)	(0.209)	(0.163)
WHITE	–0.035	–0.068	0.081	–0.128	–0.013	–0.020
	(0.093)	(0.065)	(0.171)	(0.115)	(0.312)	(0.241)
AGEPREV	–0.003	–0.003	0.003	0.003	0.017*	0.018*
	(0.003)	(0.003)	(0.004)	(0.004)	(0.007)	(0.006)
AGEPREV*BLACK	0.000	0.001	–0.003	–0.003	–0.014*	–0.015*
	(0.003)	(0.003)	(0.005)	(0.004)	(0.007)	(0.006)
AGEPREV*COLOURED	0.003	0.003	–0.001	–0.001	–0.010	–0.011
	(0.003)	(0.003)	(0.005)	(0.005)	(0.007)	(0.007)
AGEPREV*WHITE	0.004	0.004	0.007	0.008	–0.001	–0.001
	(0.003)	(0.003)	(0.005)	(0.005)	(0.010)	(0.009)

Table 10: *Continued*

	Transition 1 → 2		Transition 2 → 3		Transition 3 → 4	
	Educ'n	Income quantiles	Educ'n	Income quantiles	Educ'n	Income quantiles
AGEMOTH	0.005*	0.005*	0.000	–0.001	–0.002	–0.002
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)
AGEMOTH*BLACK	–0.001	–0.001	0.002	0.002	0.002	0.003
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)
AGEMOTH*COLOURED	–0.003*	–0.003*	0.000	0.001	0.000	0.000
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)
AGEMOTH*WHITE	0.000	0.000	–0.001	–0.001	0.001	0.000
	(0.001)	(0.001)	(0.003)	(0.003)	(0.006)	(0.005)
RURAL	1.183	1.061	–0.043	–0.041	0.002	–0.003
	(45.903)	(26.923)	(0.037)	(0.036)	(0.055)	(0.053)
RURAL*BLACK	–1.200	–1.079	0.008	0.006	–0.033	–0.030
	(45.903)	(26.923)	(0.038)	(0.037)	(0.056)	(0.054)
RURAL*COLOURED	–1.210	–1.080				
	(45.903)	(26.923)				
RURAL*WHITE	–1.192	–1.069	–0.023	–0.017	–0.073	–0.076
	(45.903)	(26.923)	(0.051)	(0.051)	(0.089)	(0.083)
EDUCHD1	0.018		0.041		0.116	
	(0.060)		(0.093)		(0.123)	
EDUCHD1*BLACK	–0.019		–0.054		–0.132	
	(0.060)		(0.093)		(0.123)	
EDUCHD1*COLOURED	0.018		–0.076		–0.119	
	(0.064)		(0.098)		(0.130)	
EDUCHD1*WHITE	–0.162*		–0.178		1.152	
	(0.079)		(0.144)		(72.930)	
EDUCHD2	–0.024		0.064		0.074	
	(0.055)		(0.086)		(0.117)	
EDUCHD2*BLACK	0.018		–0.073		–0.077	
	(0.056)		(0.086)		(0.117)	
EDUCHD2*COLOURED	0.043		–0.091		–0.067	
	(0.060)		(0.091)		(0.124)	
EDUCHD2*WHITE	–0.001		–0.175		–0.056	
	(0.065)		(0.127)		(0.196)	
EDUCHD3	0.016		0.063		0.081	
	(0.055)		(0.085)		(0.118)	
EDUCHD3*BLACK	–0.015		–0.070		–0.094	
	(0.055)		(0.086)		(0.119)	
EDUCHD3*COLOURED	–0.041		–0.094		–0.098	
	(0.061)		(0.093)		(0.130)	

Table 10: *Continued*

	Transition 1 → 2		Transition 2 → 3		Transition 3 → 4	
	Educ'n	Income quantiles	Educ'n	Income quantiles	Educ'n	Income quantiles
EDUCHD3*WHITE	–0.047		–0.191		–0.016	
	(0.063)		(0.125)		(0.192)	
INC1		0.016		0.014		0.017
		(0.026)		(0.039)		(0.061)
INC1*BLACK		–0.005		–0.010		0.005
		(0.027)		(0.039)		(0.061)
INC1*COLOURED		–0.005		0.008		–0.056
		(0.031)		(0.044)		(0.067)
INC1*WHITE		–0.021		0.052		–0.124
		(0.030)		(0.048)		(0.081)
INC2		0.041		–0.037		–0.070
		(0.038)		(0.048)		(0.062)
INC2*BLACK		–0.038		0.043		0.074
		(0.039)		(0.048)		(0.062)
INC2*COLOURED		–0.031		0.013		0.002
		(0.042)		(0.052)		(0.069)
INC2*WHITE		–0.069		0.078		0.085
		(0.050)		(0.073)		(0.110)
CONSTANT	1.182*	1.187*	1.207*	1.285*	0.913*	1.001*
	(0.081)	(0.056)	(0.126)	(0.089)	(0.191)	(0.149)
κ	0.613*	0.609*	0.626*	0.620*	0.656*	0.647*
	(0.032)	(0.032)	(0.041)	(0.040)	(0.052)	(0.052)
σ	0.149*	0.149*	0.152*	0.152*	0.146*	0.146*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Number of observations	8839	8839	5763	5763	3528	3528
Number censored	3076	3076	2235	2235	1460	1460
Log likelihood	1261.1805	1254.2308	377.4860	378.4441	171.5568	176.6765

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Educ'n, education of household head.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution.

The results for the Indian households (Table 11) are similar to those presented in Table 8. The coefficient estimates of SEXPREV and TOTMAL are always positive, though not always statistically

Table 11: *Robustness — Indian Households Only*

	Transition 1 → 2		Transition 2 → 3		Transition 3 → 4	
	Educ'n	Income quantiles	Educ'n	Income quantiles	Educ'n	Income quantiles
SEXPREV	0.054*	0.051*	0.082**	0.078	0.075	0.078
	(0.021)	(0.021)	(0.047)	(0.049)	(0.048)	(0.051)
TOTMAL			0.080*	0.078*	0.032	0.037
			(0.029)	(0.029)	(0.026)	(0.026)
DIFFSEX			–0.044	–0.046	–0.020	–0.019
			(0.029)	(0.029)	(0.044)	(0.044)
AGEPREV	–0.005**	–0.004	0.001	0.001	0.017*	0.016*
	(0.003)	(0.003)	(0.004)	(0.004)	(0.007)	(0.006)
AGEMOTH	0.005*	0.005*	0.001	0.001	–0.002	–0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.003)
EDUCHD1	0.025		0.061		0.143	
	(0.060)		(0.083)		(0.126)	
EDUCHD2	–0.012		0.081		0.104	
	(0.056)		(0.076)		(0.122)	
EDUCHD3	0.030		0.084		0.081	
	(0.055)		(0.075)		(0.122)	
INC1		0.000		0.000		0.000
		(0.000)		(0.000)		(0.000)
INC2		0.000		0.000		0.000
		(0.000)		(0.000)		(0.000)
CONSTANT	1.221*	1.259*	1.244*	1.325*	0.987*	1.036
	(0.084)	(0.064)	(0.104)	(0.090)	(0.164)	(0.155)
α	7.275*	7.135*	8.531*	8.467*	8.787*	8.553
	(0.408)	(0.392)	(0.620)	(0.611)	(1.058)	(1.002)
Number of observations	262	262	187	187	93	93
Number censored	75	75	94	94	56	56
Log likelihood	31.1552	29.5727	20.6865	20.3681	–5.3831	–6.1528

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Educ'n, education of household head.
*Significant at the 95% level; **significant at the 90% level.
α: Parameter of the Weibull distribution.

significant. Therefore, irrespective of the number of existing children, the duration is higher following the birth of a son. Further, the greater the number of existing sons, the greater is the duration between successive children. The coefficient estimate of DIFFSEX is always negative, but is never statistically significant. The results on son preference for Indian households are therefore fairly robust.

5. Concluding Comments

This paper examines the son preference hypothesis using individual level unit record data from South Africa. While the preference for sons is observed to exist in many developing countries, particularly in Asia, evidence for South Africa is scarce. The presence of four distinct races makes it an interesting country to study and the presence of a strong Asian (Indian) community within the country allows researchers to make comparisons with the Asian countries. We have used an accelerated hazard model to determine the duration between successive births and our estimation results indicate that there are significant differences among the races in the extent of preference for sons over daughters. In particular, son preference is observed to exist primarily for the Indian community in South Africa. For the Indian households, regardless of the number of existing sons, the duration is higher after a son. Additionally, an increase in the number of sons also significantly increases the duration between the second and the third children. For the non-Indian households, we do not find any evidence of son preference.

What are the possible reasons for this observed preference for sons amongst the Indian households and not among households belonging to the other races? Old age security could be an important factor, because the universal social pension programme that is now available to almost all elderly South Africans (subject to a generous means test) was extended to the non-White races only in the 1980s. Prior to this, elderly parents had to depend on their children for old age support. Given the structure of the Indian society, it is not surprising that parents in Indian families depended on their sons more than their daughters. However, we do not want to emphasise the old age security argument in explaining son preference among Indians because the same argument should hold for the Blacks. We therefore need to look at other explanations. While we cannot elaborate without additional data, it is likely that social customs and traditions play an important role. Of particular importance is the *kutum*, which as we have already noted is paternalistic and male dominated. Moreover, there is ethnographic evidence that relations between men and women vary significantly between races. For example, traditional marriage agreements for the Zulus involve a payment made by the groom and his family to the family of the bride (*lobola*). In contrast, for the Indians the payment is the other way around, with payment flowing from the bride's family to the groom and his family (*dowry*). This might partly explain the differences in parental preferences across the races.

References

Ali Khan, M. and I. Sirageldin (1977) 'Son Preference and the Demand for Additional Children in Pakistan', *Demography*, 14 (4): 481–95.

Aly, H.S. and M.P. Shields (1991) 'Son Preference and Contraception in Egypt', *Economic Development and Cultural Change*, 39 (2): 353–70.

Arnold, F., M.K. Choe and T.K. Roy (1998) 'Son Preference, the Family Building Process and Child Mortality in India', *Population Studies*, 52: 301–15.

Carter, M.R. and J. May (1999) 'Poverty Livelihood and Class in Rural South Africa', *World Development*, 27 (1): 1–20.

Chowdhury, M.K. and R. Bairagi (1990) 'Son Preference and Fertility in Bangladesh', *Population and Development Review*, 16 (4): 749–57.

Dasgupta, M. (1987) 'Selective Discrimination against Female Children in Rural Punjab', *Population and Development Review*, 13.

Gangadharan, L. and P. Maitra (2000) 'Does Child Mortality Reflect Gender Bias? Evidence from Pakistan', *Indian Economic Review*, XXXV (2): 113–31.

Gangadharan, L. and P. Maitra (2001) 'Two Aspects of Fertility Behaviour in South Africa', *Economic Development and Cultural Change*, 50 (1): 183–200.

Garg, A. and J. Morduch (1998) 'Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana', *Journal of Population Economics*, 11 (4): 471–93.

Haughton, J. and D. Haughton (1998) 'Are Simple Tests of Son Preference Useful? An Evaluation using Data from Vietnam', *Journal of Population Economics*, 11 (4): 495–16.

Johansson, S. and O. Nygren (1991) 'The Missing Girls of India', *Population and Development Review*, 17 (1): 35–51.

Jooma, A. (1991) *Migrancy After Influx Control*, Johannesburg: South African Institute for Race Relations.

Kishor, S. (1993) '"May God Give Sons to All": Gender and Child Mortality in India', *American Sociological Review*, 58: 247–65.

Kuper, H. (1960) *Indian People of Natal*, Durban: Natal University Press.
Larsen, U., W. Chung and M. Dasgupta (1998) 'Fertility and Son Preference in Korea', *Population Studies*, 52: 317–25.
Lemon, A. (1987) *Apartheid in Transition*, Vermont, Gower.
Leung, S.F. (1988) 'On Tests for Sex Preferences', *Journal of Population Economics*, 1: 95–114.
Maluccio, J., L. Haddad and J. May (2000) 'Social Capital and Household Welfare in South Africa, 1993–98', *The Journal of Development Studies*, 36 (6): 54–81.
Morduch, J. (2000) 'Sibling Rivalry in Africa', *American Economic Review Papers and Proceedings*, 90 (2): 405–9.
Quisumbing, A.R. and J. Maluccio (2000) 'Intrahousehold Allocation and Gender Relations: New Empirical Evidence', Discussion Paper, Washington DC: International Food Policy Research Institute.
Rahman, M. and J. Da Vanzo (1993) 'Gender Preference and Birth Spacing in Matlab, Bangladesh', *Demography*, 30 (3): 315–32.
Raut, L.K. (1996) 'Old-age Security and the Gender Preference Hypothesis: A Duration Analysis of Malaysian Family Life Survey Data', *Journal of Quantitative Economics*, 12 (1): 81–104.
Rose, E. (1999) 'Consumption Smoothing and Excess Female Mortality in Rural India', *The Review of Economics and Statistics*, 81 (1): 41–9.
Rosenzweig, M.R. (1986) 'Birth Spacing and Sibling Inequality: Asymmetric Information within the Family', *International Economic Review*, 27 (1): 55–76.
Rosenzweig, M.R. and T.P. Schultz (1982) 'Market Opportunities, Genetic Endowments and Intrafamily Resource Distribution: Child Survival in Rural India', *American Economic Review*, 72 (4): 803–15.
SALDRU (1994) *South Africans Rich and Poor: Baseline Economic Statistics*, Cape Town: School of Economics, University of Cape Town.
Sen, A.K. and S. Sengupta (1983) 'Malnutrition of Rural Children and Sex Bias', *Economic and Political Weekly*, 855–64.
Thomas, D. (1994) 'Like Father, Like Son; Like Mother, Like Daughter: Parental Resources and Child Height', *Journal of Human Resources*, 29 (4): 950–88.
Yach, D. (1988) 'Infant Mortality Rates in Urban Areas of South Africa, 1981–1985', *South African Medical Journal*, 73 (4): 232–34.

Appendix: Alternative Specifications

Table A1: *Alternative Specifications, No Race Interactions: Transition 1 → 2*

	Specification 1	Specification 2	Specification 3
SEXPREV		–0.003	–0.003
		(0.004)	(0.004)
BLACK	0.057*	0.057*	0.060*
	(0.014)	(0.014)	(0.014)
COLOURED	0.010	0.009	0.013
	(0.016)	(0.016)	(0.016)
WHITE	–0.004	–0.005	–0.010
	(0.015)	(0.015)	(0.015)
AGEPREV	–0.002*	–0.002*	–0.002*
	(0.000)	(0.000)	(0.000)
AGEMOTH	0.004*	0.004*	0.004*
	(0.000)	(0.000)	(0.000)
RURAL	–0.015*	–0.015*	–0.013*
	(0.006)	(0.006)	(0.006)
EDUCM1			–0.009
			(0.006)
EDUCM2			0.000
			(0.006)
EDUCM3			0.015*
			(0.007)
CONSTANT	1.244*	1.246*	1.240*
	(0.016)	(0.016)	(0.018)
κ	0.614*	0.614*	0.613*
	(0.032)	(0.032)	(0.032)
σ	0.149*	0.149*	0.149*
	(0.002)	(0.002)	(0.002)
	Specification 4	Specification 5	Specification 6
SEXPREV		0.048*	0.048*
		(0.021)	(0.021)
SEXPREV*BLACK		–0.055*	–0.055*
		(0.022)	(0.022)
SEXPREV*COLOURED		–0.048**	–0.047**
		(0.025)	(0.025)
SEXPREV*WHITE		–0.047**	–0.049*
		(0.025)	(0.025)
BLACK	0.100**	0.123*	0.110
	(0.056)	(0.056)	(0.080)

Table A1: *Continued*

	Specification 4	Specification 5	Specification 6
COLOURED	0.069	0.090	0.091
	(0.061)	(0.062)	(0.090)
WHITE	–0.090	–0.069	–0.104
	(0.064)	(0.065)	(0.088)
AGEPREV	–0.003	–0.003	–0.003
	(0.003)	(0.003)	(0.003)
AGEPREV*BLACK	0.001	0.000	0.000
	(0.003)	(0.003)	(0.003)
AGEPREV*COLOURED	0.003	0.003	f0.002
	(0.003)	(0.003)	(0.003)
AGEPREV*WHITE	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)
AGEMOTH	0.005*	0.005*	0.004*
	(0.001)	(0.001)	(0.001)
AGEMOTH*BLACK	–0.001	–0.001	0.000
	(0.001)	(0.001)	(0.001)
AGEMOTH*COLOURED	–0.003*	–0.003*	–0.003**
	(0.001)	f(0.001)	(0.002)
AGEMOTH*WHITE	0.000	0.000	0.000
	(0.001)	(0.001)	(0.002)
RURAL	1.027	1.056	1.111
	(26.439)	(27.423)	(36.485)
RURAL*BLACK	–1.044	–1.072	–1.125
	(26.439)	(27.423)	(36.485)
RURAL*COLOURED	–1.044	–1.072	–1.149
	(26.439)	(27.423)	(36.485)
RURAL*WHITE	–1.035	–1.063	–1.121
	(26.439)	(27.423)	(36.485)
EDUCM1			0.023
			(0.044)
EDUCM1*BLACK			–0.036
			(0.044)
EDUCM1*COLOURED			–0.003
			(0.052)
EDUCM1*WHITE			–0.007
			(0.056)
EDUCM2			–0.006
			(0.042)
EDUCM2*BLACK			0.005
			(0.042)
EDUCM2*COLOURED			–0.021
			(0.050)

Table A1: *Continued*

	Specification 4	Specification 5	Specification 6
EDUCM2*WHITE			0.053
			(0.049)
EDUCM3			–0.003
			(0.045)
EDUCM3*BLACK			0.028
			(0.045)
EDUCM3*COLOURED			0.007
			(0.056)
EDUCM3*WHITE			0.021
			(0.050)
CONSTANT	1.208*	1.187*	1.198*
	(0.055)	(0.056)	(0.079)
κ	0.606*	0.609*	0.616*
	(0.032)	(0.032)	(0.032)
σ	0.150*	0.149*	0.149*
	(0.002)	(0.002)	(0.002)

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution.

Table A2: *Alternative Specifications, No Race Interactions: Transition 2 → 3*

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5	Spec. 6
SEXPREV		–0.007			–0.012**	–0.011
		(0.005)			(0.007)	(0.007)
TOTMAL			–0.001		0.004	0.004
			(0.003)		(0.005)	(0.005)
DIFFSEX				–0.001	–0.001	–0.001
				(0.005)	(0.005)	(0.005)
BLACK	–0.019	–0.018	–0.019	–0.018	–0.018	–0.012
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
COLOURED	–0.034	–0.034	–0.034	–0.034	–0.034	–0.030
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)

Table A2: *Continued*

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5	Spec. 6
WHITE	0.010	0.010	0.010	0.010	0.010	0.002
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
AGEPREV	0.001**	0.001**	0.001**	0.001**	0.001**	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AGEMOTH	0.001*	0.001*	0.001*	0.001*	0.001*	0.002*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RURAL	–0.035*	–0.035*	–0.035*	–0.035*	–0.035*	–0.028*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
EDUCM1						–0.003
						(0.007)
EDUCM2						0.018*
						(0.008)
EDUCM3						0.035*
						(0.010)
CONSTANT	1.320*	1.324*	1.322*	1.321*	1.323*	1.296*
	(0.022)	(0.022)	(0.023)	(0.022)	(0.023)	(0.025)
κ	0.628*	0.630*	0.629*	0.628*	0.630*	0.630*
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
σ	0.153*	0.152*	0.153*	0.153*	0.152*	0.152*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

	Spec. 7	Spec. 8	Spec. 9	Spec. 10	Spec. 11	Spec. 12
SEXPREV		0.030			0.103**	0.107**
		(0.030)			(0.055)	(0.056)
SEXPREV*BLACK		–0.034			–0.099**	–0.103**
		(0.031)			(0.056)	(0.057)
SEXPREV*COLOURED		–0.075*			–0.033	–0.037
		(0.035)			(0.061)	(0.062)
SEXPREV*WHITE		–0.048			–0.102	–0.093
		(0.037)			(0.065)	(0.066)
TOTMAL			0.051*		0.099*	0.097*
			(0.020)		(0.034)	(0.034)
TOTMAL*BLACK			–0.053*		–0.099*	–0.097*
			(0.020)		(0.034)	(0.034)
TOTMAL*COLOURED			–0.060*		–0.073**	–0.072*
			(0.023)		(0.038)	(0.038)
TOTMAL*WHITE			–0.064*		–0.110*	–0.101*
			(0.024)		(0.040)	(0.041)

Table A2: *Continued*

	Spec. 7	Spec. 8	Spec. 9	Spec. 10	Spec. 11	Spec. 12
DIFFSEX				0.046	0.053	0.059**
				(0.032)	(0.034)	(0.034)
DIFFSEX*BLACK				–0.057**	–0.065**	–0.071*
				(0.032)	(0.034)	(0.035)
DIFFSEX*COLOURED				0.006	–0.004	–0.009
				(0.036)	(0.038)	(0.038)
DIFFSEX*WHITE				0.012	0.004	–0.002
				(0.039)	(0.040)	(0.041)
BLACK	–0.004	0.011	0.016	0.034	0.036	0.011
	(0.087)	(0.088)	(0.086)	(0.091)	(0.090)	(0.120)
COLOURED	–0.034	–0.001	–0.004	–0.013	–0.011	0.002
	(0.096)	(0.097)	(0.096)	(0.099)	(0.099)	(0.135)
WHITE	–0.133	–0.108	–0.099	–0.121	–0.107	–0.094
	(0.112)	(0.112)	(0.112)	(0.114)	(0.115)	(0.145)
AGEPREV	0.005	0.005	0.004	0.005	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
AGEPREV*BLACK	–0.004	–0.004	–0.004	–0.004	–0.003	–0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
AGEPREV*COLOURED	–0.002	–0.002	–0.002	–0.002	–0.001	0.000
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
AGEPREV*WHITE	0.006	0.005	0.006	0.005	0.007	0.007
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
AGEMOTH	–0.001	–0.001	–0.001	0.000	–0.001	–0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AGEMOTH*BLACK	0.002	0.002	0.003	0.002	0.002	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
AGEMOTH*COLOURED	0.001	0.001	0.002	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
AGEMOTH*WHITE	0.000	0.000	0.000	–0.001	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
RURAL	–0.053	–0.055	–0.055	–0.060**	–0.047	–0.049
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
RURAL*BLACK	0.020	0.022	0.022	0.027	0.014	0.026
	(0.037)	(0.036)	(0.037)	(0.037)	(0.037)	(0.037)
RURAL*WHITE	–0.003	0.005	0.002	0.010	–0.012	0.006
	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.052)
EDUCM1						0.062
						(0.065)
EDUCM1*BLACK						–0.063
						(0.065)

Table A2: *Continued*

	Spec. 7	Spec. 8	Spec. 9	Spec. 10	Spec. 11	Spec. 12
EDUCM1*COLOURED						–0.091
						(0.074)
EDUCM1*WHITE						–0.163**
						(0.087)
EDUCM2						0.013
						(0.063)
EDUCM2*BLACK						0.010
						(0.064)
EDUCM2*COLOURED						–0.021
						(0.073)
EDUCM2*WHITE						–0.044
						(0.078)
EDUCM3						0.029
						(0.065)
EDUCM3*BLACK						0.013
						(0.066)
EDUCM3*COLOURED						–0.063
						(0.079)
EDUCM3*WHITE						–0.039
						(0.078)
CONSTANT	1.311*	1.298*	1.292*	1.278*	1.278*	1.277*
	(0.087)	(0.087)	(0.086)	(0.090)	(0.089)	(0.119)
κ	0.633*	0.632*	0.630*	0.626*	0.623*	0.618*
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
σ	0.152*	0.152*	0.152*	0.152*	0.152*	0.153*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Spec., specification.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution.

Table A3: *Alternative Specifications, No Race Interactions: Transition 3 → 4*

	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5	Spec. 6
SEXPREV		–0.008			–0.001	–0.002
		(0.006)			(0.008)	(0.008)
TOTMAL			–0.007		–0.007	–0.006
			(0.003)		(0.004)	(0.004)
DIFFSEX				0.002	0.001	0.000
				(0.007)	(0.007)	(0.007)
BLACK	–0.012	–0.013	–0.011	–0.013	–0.012	–0.005
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
COLOURED	0.003	0.002	0.004	0.003	0.004	0.007
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
WHITE	0.050	0.049	0.051	0.050	0.051	0.044
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
AGEPREV	0.003*	0.003*	0.003*	0.003*	0.003*	0.003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AGEMOTH	0.000	0.000	0.000	0.000	0.000	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
RURAL	–0.031*	–0.031*	–0.030*	–0.031*	–0.030*	–0.024*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
EDUCM1						0.003
						(0.008)
EDUCM2						0.020*
						(0.009)
EDUCM3						0.035*
						(0.014)
CONSTANT	1.298*	1.302*	1.306*	1.297*	1.305*	1.278*
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.034)
κ	0.647*	0.645*	0.646*	0.647*	0.646*	0.649*
	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)
σ	0.147*	0.147*	0.147*	0.147*	0.147*	0.146*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)

	Spec. 7	Spec. 8	Spec. 9	Spec. 10	Spec. 11	Spec. 12
SEXPREV		0.022			0.045	0.037
		(0.045)			(0.053)	(0.054)
SEXPREV*BLACK		–0.029			–0.046	–0.039
		(0.045)			(0.053)	(0.054)
SEXPREV*COLOURED		–0.052			–0.051	–0.041
		(0.051)			(0.060)	(0.062)
SEXPREV*WHITE		0.001			–0.019	–0.018
		(0.062)			(0.073)	(0.074)

Table A3: *Continued*

	Spec. 7	Spec. 8	Spec. 9	Spec. 10	Spec. 11	Spec. 12
TOTMAL			–0.013		–0.025	–0.014
			(0.025)		(0.028)	(0.032)
TOTMAL*BLACK			0.007		0.020	0.009
			(0.025)		(0.029)	(0.032)
TOTMAL*COLOURED			–0.013		0.002	–0.012
			(0.028)		(0.033)	(0.036)
TOTMAL*WHITE			0.019		0.031	0.021
			(0.035)		(0.038)	(0.041)
DIFFSEX				0.019	0.013	–0.008
				(0.049)	(0.049)	(0.053)
DIFFSEX*BLACK				–0.021	–0.015	0.005
				(0.049)	(0.049)	(0.054)
DIFFSEX*COLOURED				0.011	0.017	0.038
				(0.055)	(0.055)	(0.059)
DIFFSEX*WHITE				0.071	0.081	0.100
				(0.066)	(0.067)	(0.070)
BLACK	0.294*	0.308*	0.291*	0.311*	0.314*	0.386*
	(0.145)	(0.146)	(0.148)	(0.150)	(0.152)	(0.183)
COLOURED	0.292**	0.322*	0.320*	0.294**	0.328*	0.429*
	(0.158)	(0.160)	(0.162)	(0.163)	(0.165)	(0.205)
WHITE	0.055	0.049	0.028	–0.013	–0.066	0.066
	(0.226)	(0.228)	(0.237)	(0.231)	(0.247)	(0.283)
AGEPREV	0.016*	0.016*	0.017*	0.016*	0.017*	0.016*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
AGEPREV*BLACK	–0.013*	–0.013*	–0.014*	–0.013*	–0.014*	–0.014*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
AGEPREV*COLOURED	–0.009	–0.010	–0.009	–0.010	–0.010	–0.009
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
AGEPREV*WHITE	–0.002	–0.001	–0.002	0.000	0.001	–0.001
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
AGEMOTH	–0.001	–0.001	–0.001	–0.001	–0.002	0.000
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
AGEMOTH*BLACK	0.002	0.002	0.002	0.002	0.002	0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
AGEMOTH*COLOURED	0.000	0.000	0.000	0.000	0.000	–0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
AGEMOTH*WHITE	0.001	0.001	0.001	0.000	0.000	–0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
RURAL	–0.018	–0.071	–0.071	–0.016	–0.073	–0.078
	(0.052)	(0.063)	(0.063)	(0.052)	(0.063)	(0.064)

Table A3: *Continued*

	Spec. 7	Spec. 8	Spec. 9	Spec. 10	Spec. 11	Spec. 12
RURAL*BLACK	–0.013	0.040	0.041	–0.016	0.043	0.054
	(0.053)	(0.064)	(0.064)	(0.053)	(0.064)	(0.065)
RURAL*WHITE	–0.051	0.056	0.065	–0.054	0.070	0.078
	(0.082)	(0.082)	(0.083)	(0.082)	(0.083)	(0.084)
EDUCM1						–0.034
						(0.097)
EDUCM1*BLACK						0.037
						(0.097)
EDUCM1*COLOURED						0.017
						(0.109)
EDUCM1*WHITE						0.061
						(0.154)
EDUCM2						0.052
						(0.101)
EDUCM2*BLACK						–0.032
						(0.102)
EDUCM2*COLOURED						–0.045
						(0.114)
EDUCM2*WHITE						–0.088
						(0.141)
EDUCM3						0.053
						(0.108)
EDUCM3*BLACK						–0.017
						(0.109)
EDUCM3*COLOURED						–0.118
						(0.125)
EDUCM3*WHITE						–0.025
						(0.145)
CONSTANT	0.997*	0.985*	1.008*	0.983*	0.987*	0.895*
	(0.145)	(0.146)	(0.147)	(0.149)	(0.151)	(0.182)
κ	0.648*	0.646*	0.648*	0.651*	0.653*	0.657*
	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.053)
σ	0.146*	0.146*	0.146*	0.146*	0.146*	0.145*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)

Robust standard errors in parentheses. Regressions control for pre-1994 province dummies. Spec. specification.
*Significant at the 95% level; **significant at the 90% level.
κ, σ: Parameters of the gamma distribution.