

# Testing for Son Preference in South Africa\*

Lata Gangadharan

Department of Economics

University of Melbourne

Parkville, VIC 3052, Australia

E-mail: gangadha@cupid.econ.unimelb.edu.au

Pushkar Maitra

Department of Economics

University of Sydney

Sydney NSW 2006, Australia

E-mail: pushkarm@econ.usyd.edu.au

November 1999

---

\*We would like to thank Siwan Anderson, Bob Bartels, Monojit Chatterji, Patrick Francois, Elizabeth Savage, Don Wright, participants in the Friday workshop at the Department of Economics, Sydney University, participants at the Data Modelling workshop in the Department of Econometrics, Sydney University and participants at the 28<sup>th</sup> Conference of Economists held at La Trobe University for comments and discussions. We are however responsible for all remaining errors.

# Abstract

Evidence from most developing countries suggests that parents have a preference for sons over daughters. This is known 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 prefer 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 either no son preference or in some cases a weak preference for daughters. Our results appear to refute the usual arguments for son preference (including support for elderly parents and contribution to household income) and instead suggest the importance of religious beliefs and social customs (dowry system, lineage, familial and kinship ties etc.) in directing parental preferences.

JEL Classification: J16, D12, C41, C24.

Keywords: Son Preference, Duration between births, South Africa.

# 1 Introduction

The gender preference hypothesis postulates that parents exhibit preferences for having children of a particular gender. In most 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 income and provide old-age parental support. Daughters on the other hand are married off 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. Leung (1988) and Raut (1996) find evidence of preference for sons among the Chinese but not the Malays in Malaysia. Larsen, Chung & Dasgupta (1998) using data from Korea find that women who have a son are less likely to have another child and if they do proceed to have another child, take longer to conceive the subsequent child. Khan & Sirageldin (1977) find that son preference affects the actual childbearing intentions of parents in Pakistan and, using data from Egypt, Aly & Shields (1993) find that the probability of a woman using contraception increases with the number of existing male children. Das (1987) and Arnold, Kim & Roy (1998) using data from India find that son preference fundamentally affects demographic behaviour and has significant impact on fertility and child mortality.

This paper tests for son preference in South Africa. 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.<sup>1</sup> Moreover the presence of a sizeable Indian community in South Africa

---

<sup>1</sup>During the apartheid era all South Africans were classified into one of the four following race categories: Black

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 \$3000 [see Carter & 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) 3347.73 (nearly \$1000) in 1993 - 94. An evidence of son preference within the Indian community could refute the usual arguments for preferring sons over daughters (old age security, addition to household income) and would instead support arguments like religious beliefs and social customs (dowry system, lineage, familial and kinship ties) in directing parental preferences towards sons.

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. Parents care about the quality of their children and the quality of a particular child depends on its birth order, on the age of the parents when born, the interval between its birth and prior and successive births, its own endowments and also on a set of parental, household and sibling endowments. We use an accelerated hazard model to estimate the duration between successive births for a set of women in South Africa. We examine the duration between child 1 and 2 (Transition  $1 \rightarrow 2$ ), duration between child 2 and 3 (Transition  $2 \rightarrow 3$ ) and the duration between child 3 and 4 (Transition  $3 \rightarrow 4$ ). If there is indeed a preference for sons over daughters we expect to observe the following:

1. An increased duration following the birth of a son.

2. The higher the number of existing sons, the greater the duration between successive children.

---

(African), Coloured (Mixed race), Indian (Asian) and White (Caucasian). To maintain consistency with the data and the existing literature we will stick to this categorisation.

Our results show that the son preference hypothesis is supported only for the Indian households. For the Indian households, irrespective of the transition, the duration is higher after the birth of a son. Additionally, an increase in the number of existing sons increases the duration between child 2 and child 3. For the rest of the population, there is either no son preference or in some cases a weak preference for daughters.

Our paper contributes to the literature on gender preferences in the following ways.

1. We use an accelerated hazard analysis of the duration between successive births. Most analyses of duration between children use a proportional hazard model. This model leads to possible mis-specification because it does not take into account the effect of time on the hazard of having another child. The notable exception here is Raut (1996). However our econometric analysis extends that of Raut's in that we use a Gamma distribution which is the most general distribution and encompasses the log normal and the Weibull distribution that are used by Raut.
2. This paper, instead of using a simple reduced form methodology, first shows how preference for sons can arise from a set up where parents care about the quality of their children. The theoretical model yields testable implication which we then test using individual level unit record data from South Africa.
3. This paper is possibly the first to study son preference among South African households. The presence of four distinct races makes South Africa an interesting country to study. As already argued, we find that the Indian households behave very differently from the rest of the population and exhibit a strong preference for sons.

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 of Indians out of Natal.<sup>2</sup> 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. The extended family (or the *kutum*) is a very important aspect of Indian life in South Africa and it is this extended family that has overtime insulated Indians from the oppressions of the South African society under apartheid. The Indian *kutum* is extremely male dominated and consists of all families who can be traced to a common paternal grand father or the grand father'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. [See Lemon (1987) for a survey of Indians in South Africa]. The extended family and paternal society makes Indians different from the rest of the South African population. On the other hand, it is surprising that even though the Indians in South Africa have a history going back to 1860's, 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. 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 had been born outside of South Africa. Actually of the 890,000 South African Indians, more than 70% are descendants of the indentured immigrants. Today Indians in South Africa form a rich, urban upper middle-class, but their preferences regarding the gender of their children are very

---

<sup>2</sup>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.

similar to Indians in the sub-continent.

The implications of our results are quite disappointing. Economic arguments to explain son preference include the old age security motive for having sons or the fact that in rural households sons can provide an additional hand in the farms. One expects that as the economy develops, the availability of alternative means of old age pension would reduce the old age security motive for preferring sons and the availability of improved farming technology would reduce the demand for child labour on farms. In South Africa, no Indian household resides in rural communities and there is a means-tested social pension program that reduces the dependence on children for old age support. Hence the traditional arguments for son preference do not appear to exist. One has to therefore acknowledge the role of customs and traditions (including the paternalistic nature of the extended family) in directing parental preferences regarding the gender composition of their children.

The rest of the paper is organised as follows. Section 2 describes the methodology used in the paper - it presents the simple theoretical framework and the testable implications of the son preference hypothesis. In Section 3, we describe the estimation technique used. Section 4 describes the data and selected descriptive statistics. Section 5 discusses the results and Section 6 examines the robustness of the results. Section 7 concludes.

## 2 Methodology

The decision horizon is divided into two equally spaced periods in each of which a child is to be born. In each period ( $i$ ) a household (comprising of a couple) has to choose between having a child early ( $S_i = 0$ ) and delaying the child ( $S_i = 1$ ), and the decision on when to have a child (early or late) is based on the information set  $\Omega$  the household has at the beginning of the period.

Households maximise an intertemporally separable utility function defined over child quality ( $Q$ ) and consumption good ( $X$ ) and the optimisation problem can be written as follows:

$$\text{Max}_{S_t} \left[ \sum_{\tau=t}^T \delta^{\tau-t} U(Q_\tau, X_\tau) \right], t = 1, 2 \quad (1)$$

subject to the per-period budget constraint that must be satisfied in each period

$$Y_t = W_t S_t + X_t \quad (2)$$

where  $Y_t$  = income and  $W_t$  = cost of birth postponement.

The quality  $Q$  (education, earnings, health etc.) of a child born to household  $j$  depends on its birth order  $i$ , the age of its parents when born  $S_i$ , the interval between its birth and prior and subsequent births, endowment of previous children and child specific resources  $Z_{ij}$ . So the quality of child  $i$  in household  $j$  (who is neither the first nor the last among  $n$  children) can then be written as:

$$Q_{ij} = \gamma_a S_{ij} + \gamma_p (S_{ij} - S_{i-1j}) + \gamma_s (S_{i+1j} - S_{ij}) + \gamma_h Z_{ij} + \mu_j + \eta_{ij} + \sum_{l=1}^{i-1} \gamma_{kl} \eta_{lj} \quad (3)$$

For the first child in household  $j$

$$Q_{1j} = \gamma_a S_{1j} + \gamma_s (S_{2j} - S_{1j}) + \gamma_h Z_{1j} + \mu_j + \eta_{1j} \quad (4)$$

and for the last child in household  $j$

$$Q_{nj} = \gamma_a S_{nj} + \gamma_p (S_{nj} - S_{n-1j}) + \gamma_h Z_{nj} + \mu_j + \eta_{nj} + \sum_{l=1}^{n-1} \gamma_{kl} \eta_{lj} \quad (5)$$



Here

$\gamma_a$  = Parental Age Effect

$\gamma_p$  = Prior Interval Effect

$\gamma_s$  = Subsequent Interval Effect

$\gamma_{kl}$  = Endowment Effect of Previous Children

$\gamma_h$  = Child Specific Resource Effect

$\mu$  = Household Endowment

$\eta_{ij}$  =  $i^{th}$  child's specific quality endowment

$$E(\mu_j \mu_k) = \sigma_\mu^2 (j = k); 0 (j \neq k)$$

$$E(\eta_{ij}, \eta_{kl}) = \sigma_\eta^2 (i = k, j = l); 0 (i \neq k, j \neq l)$$

$$E(\eta_{ij}) = E(\mu_j) = 0$$

The methodology that we use closely follows Rosenzweig (1986). However in that paper the focus was primarily on the effect of duration on child quality - including physical and mental development of children, measured by weight/height ratio and the intelligence quotient. In this paper we use the framework to predict the effect of parental, household and sibling characteristics on the duration between successive children.

Equations (3) - (5) therefore define a set of “child quality functions” that examine the effect of the timing of births and other resources to child quality. We impose the following restrictions on the parameters of the quality functions:

1. Child quality depends only on timing, spacing and household endowment and we assume

that the child-specific resource effect is zero i.e.,  $\gamma_h = 0$ .

2. The parental age effect is negative i.e.,  $\gamma_a < 0$  - the child quality is lower if the child is born to older parents. This is not implausible given that it is generally argued that the probability of the child being disease prone increases with the age of the mother.
3. The prior and the subsequent interval effects are both positive  $\gamma_p > 0, \gamma_s > 0$ . The greater the duration between successive births the greater the amount of resources that the parents can provide for any particular child, and the higher is the quality of the child.
4. The effect of an increase in the endowment of previous children is to reduce the quality of the current child i.e.,  $\gamma_{kl} < 0$ . This is also not an implausible assumption because parents might be willing to invest more (time and resources) on a more endowed child adversely affecting the quality of successive children.

Parents know the household quality endowment but do not know the child quality endowment in advance of child bearing. We can therefore write the information set for each parent as follows. At the beginning of period 2,  $\Omega_2 = \{\mu_j, \eta_{1j}, S_{1j}; \gamma\}$  i.e., parents know the household quality endowment, the endowments of the previous birth, their prior timing decisions and the child quality functions (defined in (3) - (5)). Likewise the information set at the beginning of period 1 is  $\Omega_1 = \{\mu_j; \gamma\}$ .

It is difficult to derive the solution to the parent's dynamic programming problem, defined by equation (1) and (2), given the child quality functions (equations (3) - (5)). It is however possible to obtain the comparative static effects. At the beginning of period 2 parents compare the utility of close spacing of child one and child two ( $S_2 = 0$ ) to that of wide spacing ( $S_2 = 1$ ), given the

information set  $\Omega_2$ . So we can write

$$J_2 = E_2 [U_2 | S_2 = 1; \Omega_2] - E_2 [U_2 | S_2 = 0; \Omega_2] \quad (6)$$

When  $J_2 > 0$  the second birth is postponed and when  $J_2 < 0$  the post birth interval for child one is shortened. For analytical convenience let us consider the instantaneous utility function to be linear quadratic in its arguments, so that

$$U_i = \alpha_1 Q_i - \alpha_2 Q_i^2 + \beta_1 X_i - \beta_2 X_i^2; i = 1, 2 \quad (7)$$

The main reason for using a linear-quadratic utility function is that it is analytically simple.<sup>3</sup> We can now derive the effect of an increase in the family endowment on the value of  $J_2$  and thus on the duration between births one and two as  $\frac{\partial J_2}{\partial \mu}$ .

$$\begin{aligned} J_2 &= E_2 \left[ \{ \alpha_1 Q_2 - \alpha_2 (Q_2)^2 + \beta_1 X_2 - \beta_2 (X_2)^2 \} | S_2 = 1; \Omega_2 \right] \\ &\quad - E_2 \left[ \{ \alpha_1 Q_2 - \alpha_2 (Q_2)^2 + \beta_1 X_2 - \beta_2 (X_2)^2 \} | S_2 = 0; \Omega_2 \right] \\ &= E_2 \left[ \begin{array}{c} [ \alpha_1 \{ \gamma_a S_2 + \gamma_p (S_2 - S_1) + \mu + \eta_2 + \gamma_{k1} \eta_1 \} \\ - \alpha_2 \{ \gamma_a S_2 + \gamma_p (S_2 - S_1) + \mu + \eta_2 + \gamma_{k1} \eta_1 \}^2 \\ + \beta_1 X_2 - \beta_2 (X_2)^2 ] | S_2 = 1; \Omega_2 \end{array} \right] \\ &\quad - E_2 \left[ \begin{array}{c} [ \alpha_1 \{ \gamma_a S_2 + \gamma_p (S_2 - S_1) + \mu + \eta_2 + \gamma_{k1} \eta_1 \} \\ - \alpha_2 \{ \gamma_a S_2 + \gamma_p (S_2 - S_1) + \mu + \eta_2 + \gamma_{k1} \eta_1 \}^2 \\ + \beta_1 X_2 - \beta_2 (X_2)^2 ] | S_2 = 0; \Omega_2 \end{array} \right] \end{aligned}$$

---

<sup>3</sup>One can use different functional forms for  $U_i$  at the cost of messier algebra.

This gives

$$\begin{aligned}
\frac{\partial J_2}{\partial \mu} &= E_2 [\alpha_1 - 2\alpha_2 \{\gamma_a + \gamma_p (1 - S_1) + \mu + \eta_2 + \gamma_{k1}\eta_1\}] \\
&\quad - E_2 [\alpha_1 - 2\alpha_2 \{\gamma_p (-S_1) + \mu + \eta_2 + \gamma_{k1}\eta_1\}] \\
\Rightarrow \frac{\partial J_2}{\partial \mu} &= -2\alpha_2 (\gamma_a + \gamma_p)
\end{aligned} \tag{8}$$

Now since  $\alpha_2 > 0$ , if the net effect of postponing the birth is to raise child quality for all births ( $\gamma_a + \gamma_p > 0$ ), then the more endowed households will tend to space the births more closely.

Similarly the effect of the endowment of the first child on the duration between child 1 and child 2 is  $\frac{\partial J_2}{\partial \eta_1}$  and

$$\begin{aligned}
\frac{\partial J_2}{\partial \eta_1} &= E_2 [\alpha_1 \gamma_{k1} - 2\alpha_2 \gamma_{k1} \{\gamma_a + \gamma_p (1 - S_1) + \mu + \gamma_{k1}\eta_1\}] \\
&\quad - E_2 [\alpha_1 \gamma_{k1} - 2\alpha_2 \gamma_{k1} \{\gamma_p (-S_1) + \mu + \gamma_{k1}\eta_1\}] \\
\Rightarrow \frac{\partial J_2}{\partial \eta_1} &= -2\alpha_2 \gamma_{k1} (\gamma_a + \gamma_p)
\end{aligned} \tag{9}$$

Now  $\gamma_{k1} < 0$  (the endowment effect of child one on the quality of the second child is negative), therefore  $\frac{\partial J_2}{\partial \eta_1} > 0$ . This implies that an increase in the endowment of child one increases the spacing between child one and child two.<sup>4</sup>

The timing of the first child - whether to delay the first child (choose  $S_1 = 1$ ) or have the first child early (choose  $S_1 = 0$ ) - depends on the probability distribution of the endowments of the first and the second child and the optimal responses for the second period. At the beginning of period 1, parents must compare the expected future utilities associated with the different spacing

---

<sup>4</sup>Once again we assume that  $(\gamma_a + \gamma_p > 0)$

alternatives, i.e., obtain the sign of

$$\begin{aligned}
J_1 &= E_1 [U_1|S_1 = 1; \Omega_1] - E_1 [U_1|S_1 = 0; \Omega_1] \\
&= E_1 [U_1|S_1 = 1] + E_1 \{ \text{Max} \langle E_2 (U_2|S_2 = 1, S_1 = 1), E_2 (U_2|S_2 = 0, S_1 = 1) \rangle \} \\
&\quad - E_1 [U_1|S_1 = 0] + E_1 \{ \text{Max} \langle E_2 (U_2|S_2 = 1, S_1 = 0), E_2 (U_2|S_2 = 0, S_1 = 0) \rangle \}
\end{aligned}$$

One can obtain the comparative static effects.

The results can be generalised as follows:

1. The more endowed households will tend to space births more closely.
2. If the endowment of previous children increase, the spacing between children will increase.

The problem with estimating the effect of household endowment and child endowment on the duration between successive births is that both of these are unobservable to the researcher - they are private information to the household. We therefore use observed household income as the measure of household endowment and the sex of the child as the measure of the child quality endowment.<sup>5</sup> The endowment of a child is high if the child is male and the endowment is low if the child is female. Using sex of the child as a measure of child quality endowment helps us to examine the gender preferences of parents in an empirical framework. Also it is seen that parents in many societies regard their male children to have a higher endowment compared to daughters, because often the expected return from sons is greater than the expected returns from daughters. Moreover in many societies norms dictate that sons take care of their elderly parents and it is

---

<sup>5</sup>The use of household income as a measure of household quality endowment could however lead to endogeneity problems - decisions regarding duration between births might affect labour supply decisions of women and hence affect household income [see for example Angrist & Evans (1998)]. We later consider alternative measures of household quality endowment.

sons who “carry on the family name” or “light the funeral pyre”. An alternative measure of child quality endowment could be child intelligence, however data on child intelligence is private information to the parents unlike the sex of the child which is observable to the researcher.

### 3 Estimation Technique

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  $u \in U$  where  $U$  is the set of all feasible strategies. Now  $T$  will depend on a number of factors not all of which are observable to the econometrician (for example the biological endowment of the woman). Let  $\eta$  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

$$h(t|u, \eta) \equiv \text{Probability that the event } T \text{ occurs in the time interval } (t, t + dt), \text{ given that it has not occurred until } t \text{ and given the value of the individual specific unobserved heterogeneity } (\eta) \text{ and the actual strategy followed } (u).$$

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

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

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

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

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

$$h(t|X, \eta) = \lambda_0(t) e^{X'\beta} \quad (10)$$

The most general parametrisation of the baseline hazard function is the Gamma distribution. In this case the density function is

$$f(t) = \begin{cases} \frac{|\kappa|}{\Gamma(\frac{|\kappa|}{2})} (\kappa^{-2})^{\kappa-2} \exp[\kappa^{-2}(\kappa z - e^{\kappa z})], & \text{if } \kappa \neq 0 \\ \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{z^2}{2}\right), & \text{if } \kappa = 0 \end{cases}. \quad (11)$$

and  $z = \frac{\ln t - \lambda}{\sigma}$ . We parametrize  $\lambda_j = X_j\beta$  and the parameters  $\kappa$  and  $\sigma$  are estimated using the data. Note that the hazard function for the gamma distribution is extremely flexible allowing for a large number of possible shapes including as special cases the Weibull distribution where  $\kappa = 1$ , the exponential distribution where  $\kappa = 1, \sigma = 1$  and the lognormal distribution where  $\kappa = 0$ . 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 therefore is non-monotonic and the gamma distribution has the flexibility to model this non-monotonic relationship.

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 specially in analysing issues like the duration between successive births where the hazard function is likely to depend on time. To explicitly account for the effect of time, we need to consider the accelerated hazard model. For the accelerated hazard model the parametrisation is done at a different level. Let the random variable  $T_0$  denote the duration of an event conditional on the values of all the co-variates equal to zero.  $T_0$  can therefore be viewed as the duration of the event for a woman drawn randomly from the population of women who do not control their fertility behaviour. The hazard rate of  $T_0$  is given by  $\lambda_0(t)$ . So here  $\lambda_0(t)$  denotes the natural hazard rate and  $T_0$  denotes the natural or baseline duration of the event. The effect of a co-variate  $X$  in this model is therefore to scale the natural duration  $T_0$  up or down by  $e^\beta$  accordingly as  $\beta$  is positive or negative.<sup>6</sup> So the accelerated hazard model specifies:

$$T = T_0 e^{X'\beta^* + \eta}$$

or equivalently

$$\log T = \eta + X'\beta^* + \sigma\varepsilon \tag{12}$$

where  $\log T_0 \equiv \sigma\varepsilon, \sigma > 0$ ,  $\varepsilon$  has a distribution independent of  $X$  and  $\beta^*$  is the set of parameters of interest. The distribution of  $\varepsilon$  can be derived from the baseline hazard distribution  $\lambda_0(t)$  of  $T_0$ . It might however be noted that 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 are denoted by  $\beta$  and the coefficients in the accelerated hazard

---

<sup>6</sup>Compare this to the proportional hazard model where the effect of the regressor is to move the baseline hazard rate proportionately up or down.



model are denoted by  $\beta^*$ , then  $\beta^* = -\frac{\beta}{\kappa}$ . A negative coefficient  $\beta^*$  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 we could not have using least squares techniques.

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

The estimated equation is therefore given by

$$\begin{aligned}
\ln(DURAT_i) = & \beta_0 + \beta_1 COLOURED + \beta_2 INDIAN + \beta_3 WHITE \\
& + \beta_4 SEXPREV + \beta_5 IR2SEXPREV + \beta_6 IR3SEXPREV \\
& + \beta_7 IR4SEXPREV + \beta_8 TOTMAL + \beta_9 IR2TOTMAL \\
& + \beta_{10} IR3TOTMAL + \beta_{11} IRATOTMAL + \beta_{12} DIFFSEX \\
& + \beta_{13} IR2DIFFSEX + \beta_{14} IR3DIFFSEX + \beta_{15} IRADIFFSEX \\
& + \beta_{16} AGEPREV + \beta_{17} AGEMOTH + \beta_{18} RURAL \\
& + \beta_{19} INC1 + \beta_{20} IR2INC1 + \beta_{21} IR3INC1 + \beta_{22} IRAINC1 \\
& + \beta_{23} INC2 + \beta_{24} IR2INC2 + \beta_{25} IR3INC2 + \beta_{26} IRAINC2 \\
& + \sum_{p=1}^{13} \beta_{26+p} STATE_p + \varepsilon;
\end{aligned} \tag{13}$$

where  $DURAT_i$  is the duration between child  $i$  and  $i + 1$ ,  $i = 1, 2, 3$ . However if any woman stops at  $i$  then the observed duration is the entire period from the year in which she had her  $i^{th}$  child and the survey year. These are the women who are “censored”.

The explanatory variables used in the regression for each transition 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), a dummy for the sex of the previous child ( $SEXPREV = 1$  if the previous child is male, 0 otherwise) the age of the mother at previous birth ( $AGEPREV$ ), the current age of the mother ( $AGEMOTH$ ), a dummy to indicate whether the woman lives in a rural area or not ( $RURAL = 1$  if the household lives in a rural area, 0 otherwise) and two dummy variables to indicate the income category  $INC1$  and  $INC2$  ( $INC1 = 1$  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 is in the top third of the income distribution.  $SEXPREV, TOTMAL$  and  $DIFFSEX$  are the “gender preference variables”. The use of current household income to determine past household decisions could create some problems because of the nature of the data. Current household income might incorrectly measure past household endowment and hence incorrectly measure the effect of household endowment on the timing of births. To avoid this problem we use dummies ( $INC1$  and  $INC2$ ) to determine the income category of the household, which might not have changed even though actual income possibly has. We later examine the robustness of the results by considering the education level of the household head and the education level of the mother as alternative measures of household quality endowment [see section 6]. Possible racial differences are accounted for by the inclusion of a set of race dummies (to account for differences in the

intercept) and a set of interaction terms (to account for differences in the slopes). The race dummies that we include are *COLOURED*, *INDIAN* and *WHITE* (*COLOURED* = 1 if the household is Coloured, 0 otherwise; *INDIAN* = 1 if the household is Indian, 0 otherwise and *WHITE* = 1 if the household is White, 0 otherwise). The reference category is that the household is Black. We also interact the three race dummies with the gender preference variables and the income dummies. The race interaction dummies are defined as follows:

$$IR2SEXPREV = SEXPREV \otimes COLOURED$$

$$IR3SEXPREV = SEXPREV \otimes INDIAN$$

$$IR4SEXPREV = SEXPREV \otimes WHITE$$

$$IR2TOTMAL = TOTMAL \otimes COLOURED$$

$$IR3TOTMAL = TOTMAL \otimes INDIAN$$

$$IR4TOTMAL = TOTMAL \otimes WHITE$$

$$IR2DIFFSEX = DIFFSEX \otimes COLOURED$$

$$IR3DIFFSEX = DIFFSEX \otimes INDIAN$$

$$IR4DIFFSEX = DIFFSEX \otimes WHITE$$

$$IR2INC1 = INC1 \otimes COLOURED$$

$$IR3INC1 = INC1 \otimes INDIAN$$

$$IR4INC1 = INC1 \otimes WHITE$$

$$IR2INC2 = INC2 \otimes COLOURED$$

$$IR3INC2 = INC2 \otimes INDIAN$$

$$IR4INC2 = INC2 \otimes WHITE$$

Finally we include a set of province dummies ( $STATE_1, \dots, STATE_{13}$ ) to account for any unobserved heterogeneity. The reference category is that the household lives in Transvaal. Note the province dummies used in the analysis correspond to the pre-1994 provinces. They have now changed. See Table 1 for a description of all the explanatory variables used.

In the case of Transition  $1 \rightarrow 2$ ,  $SEXPREV$  refers to the sex of the first child and  $AGEPREV$  refers to the age of the mother at which she had her first child. In this case  $TOTMAL$  and  $DIFFSEX$  are not included as explanatory variables.<sup>7</sup> In the case of Transition  $2 \rightarrow 3$ ,  $TOTMAL = \{0, 1, 2\}$  and  $DIFFSEX = 1$  if either of the following sequence of children is born  $\{b, g\}$  or  $\{g, b\}$ .<sup>8</sup>  $AGEPREV$  refers to the age of the mother when she had her second child. We can define  $TOTMAL$ ,  $DIFFSEX$  and  $AGEPREV$  in similar fashion for Transition  $3 \rightarrow 4$ . In this case therefore  $TOTMAL = \{0, 1, 2, 3\}$  and  $DIFFSEX = 1$  if any of the following sequence of children is born

$$\{(b, b, g), (b, g, b), (b, g, g), (g, g, b), (g, b, g), (g, b, b)\}$$

To examine the son preference hypothesis we need to look at the coefficients of  $SEXPREV$  and  $TOTMAL$ . These two variables are related to the child-quality endowment variables in terms of our theoretical framework [see Section 2]. 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.

---

<sup>7</sup>Notice that in this case  $TOTMAL$  and  $SEXPREV$  are identical variables and  $DIFFSEX$  is not defined.

<sup>8</sup>Here  $g$  denotes a girl and  $b$  denotes a boy.

Other testable implications of the theoretical model are

1. The more endowed households will tend to space the births more closely, which in turn implies that the coefficient of *INC1* and *INC2* should be positive - households in the lower end of the income distribution scale have a higher duration between successive births.
2. The higher the age of the parent at the previous birth, the lower is the duration between children and hence the coefficient of *AGEPREV* should be negative.

## 4 Data and Descriptive Statistics

The data set used in this paper is from the South African Integrated Household Survey (SIHS) 1993 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".<sup>9</sup> The sample consists of approximately 9000 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 computed from available data. Every member of the household was asked the identification code of his/her mother and then each mother was matched with all her children. This way we were able to obtain the child bearing history of women who have at least one child who is alive at the time of the survey. However because of the way the

---

<sup>9</sup>The "homelands" were designated residential regions for the Blacks during the apartheid regime. These were autonomous states within South Africa.

data was constructed, we were able to obtain the child characteristics for only the children who are alive at the time of the survey. In particular we have no information on the children who have died.<sup>10</sup> Note that being married is not a pre-condition to bearing children and in particular more than 25% of the women in our sample are not married [see Gangadharan & Maitra (1999)].

The child-bearing history was obtained for a subset of the women surveyed. The sample used in this paper consists of 8839 women, of whom 6981 (78.97%) are Black, 756 (8.55%) are Coloured, 262 (2.96%) are Indian and 840 (9.51%) are White. In terms of sample proportions, this sample is fairly representative of the population distribution of South Africa. The average number of children is 2.63 per Black woman, 2.33 per Coloured woman, 2.24 per Indian woman and 1.86 per White woman. The Blacks are the poorest (mean household income R 963.21), followed by the Coloured (mean household income R 1767.94), Indian (mean household income R 3347.73) and White (mean household income R 25250.93).

Table 2 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

---

<sup>10</sup>This problem does not exist in some other similar surveys - for example for the Pakistan Integrated Household Survey (PIHS), conducted by the World Bank in 1991, each "ever married woman" was asked about her child bearing history and whether each child is alive or dead at the time of the survey.

of the existing children and the duration between children, however does not 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 3 show that 65.5% Black, 65.3% Coloured, 71.4% Indian and 60.6% White women go on to have a second child. Note that 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. However the proportion of women who go on to have a third child drops significantly for the Coloured, Indian and White women. Of the women who have more than one child, only 59.5% of Coloured 49.7% of Indian and 33.3% of White women go on to have a third child. We find an even larger drop in the proportion of women who go on to have a fourth child. Table 3 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 go on to have a second child is significantly higher if the first child is a boy. The proportion of women who go on to have a third child is the highest among the Blacks and the lowest among the Whites. The proportion of Coloured, Indian and White women who go on to have 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 go on to have a third child is significantly higher if both child 1 and child 2 are girls. In case of Transition 3  $\rightarrow$  4 the sex-mix of the existing children does not significantly affect the parity progression ratios.

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”. Let us consider an example. Consider a woman who is 50 years old at the time of the survey and has only one child. One could argue that this woman has stopped at one child and this is her desired number, since she is past the child bearing age. On the other hand consider a woman who is 22 years old at the time of the survey who also has one child. For this woman the observed fertility does not allow us to say anything about her desired fertility, because she has not necessarily completed her child bearing. To account for this problem, we can either examine the parity progressions for women beyond a certain age [for example Haughton & 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, taking into account the fact that some women have not completed their families (the censoring issue).

## 5 Results

Table 4 presents the accelerated hazard regression results of equation (13) for the joint estimation across races at each transition. Panel A presents the transition from the first to the second child ( $1 \rightarrow 2$ ), Panel B the transition from the second to the third child ( $2 \rightarrow 3$ ) and Panel C the transition from the third to the fourth child ( $3 \rightarrow 4$ ). A negative sign on the coefficient decreases the duration (and increases the hazard) between successive births while a positive sign increases the duration (and decreases the hazard) between successive births. Note that at each transition the estimated acceleration factor is given by  $e^\beta$ . 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.

Let us first consider Transition  $1 \rightarrow 2$ . 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. Though *SEXPREV* is not significant for the whole sample, when it is interacted with the race dummies then for the Indian case, the interaction term (*IR3SEXPREV*) is positive and significant. For Indian households therefore the duration between child 1 and child 2 is higher if the first child is male. The estimated acceleration factors are given by 1.01, 1.06 and 1.01 for Coloured, Indian and White women respectively. What this means is that controlling for the other characteristics, the effect of having a son compared to a daughter first is to leave the duration between child 1 and child 2 unaffected for Coloured and White women and significantly increase the duration (by a factor of 1.06) for Indian women. The higher the age at which a woman has her first child, the lower is the duration between child 1 and child 2 (*AGEPREV* is negative for all women). This implies that the parental age effect on the duration between births is negative. The duration between child 1 and child 2 is lower for women living in rural areas and women in lower income households wait more to have their next child (this could be a reflection of resource constraints for having additional children).

The results for the duration between child 2 and child 3 are presented in Panel B. The race dummy for Coloured households (*COLOURED*) is negative and significant implying that the Coloured households have a shorter duration between the second and the third child as compared to Black households. Once again *SEXPREV* is not significant for the entire sample, however when it is interacted with the dummy for the Indian household, then the interaction term (*IR3SEXPREV*) increases the duration. Hence for the Indian households, if the second child is a boy it increases the duration between child 2 and child 3. We also see that *IR3TOTMAL* is

positive and significant which implies that an increase in the total number of sons the woman has, significantly reduces the hazard and consequently increases the duration between child 2 and child 3 for the Indian households. Interestingly it might be noted that  $IR2SEXPREV$  is negative and significant which implies that for the Coloured households if the second child is a boy it reduces the duration between child 2 and child 3. The corresponding estimated acceleration factors are given by 0.94, 1.10 and 1.00 for Coloured, Indian and White women respectively. Ceteris paribus, the acceleration effect of the second child being a son relative to a daughter is to leave the duration between child 2 and child 3 unaffected for White women, decrease the duration (by a factor of 0.94) for Coloured women and increase the duration (by a factor of 1.10) for Indian women. For the whole sample, if child 1 and child 2 are of different sex then the duration between child 2 and child 3 decreases. However  $IR2DIFFSEX$ ,  $IR3DIFFSEX$  and  $IR4DIFFSEX$  are all positive and significant which implies that Coloured, Indian and White women wait longer than Black women for their third child as they already have one child of each gender. Women residing in rural areas have higher hazard ratios and wait less for another child.

Panel C presents the results for Transition 3  $\rightarrow$  4. The sex preference variables are not significant for the entire sample. Women in rural areas have a lower duration between child 3 and child 4 and women in lower income households wait more to have their next child. Interacting income with race however shows that White women in the lower income group and Coloured women in medium income households have a lower duration between child 3 and child 4 ( $IR4INC1$  and  $IR2INC2$  are negative and significant).

The duration analysis for the entire sample suggests that the results for the four races in South Africa are rather different. When we jointly estimate the durations across all races we are in effect imposing a restriction that the scale and shape parameters for the distribution

( $\kappa$  and  $\sigma$ ) 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  $H_0 : \kappa_1 = \kappa_2 = \kappa_3 = \kappa_4, \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$  is rejected using a standard Likelihood ratio test. Hence it is more efficient to estimate the duration between successive children separately by race. These results are presented in Table 5 (a, b and b). While the results are similar to the ones discussed above, separating the analysis by race, further strengthens the view that gender preference is significant for Indian households. Notice that for the Indian households the preferred distribution is Weibull while for the other races, the preferred distribution is Gamma. For the Indian households  $\kappa = 1$  cannot be rejected using a standard Wald test and the lowest Akaike Information Criterion is obtained by using a Weibull distribution.

Our results therefore suggest that the preference for sons exists only for the Indian households. Table 6 presents the results separately for the Indian households at the three transitions. The regression results show that the predicted duration between child 1 and child 2 is 48 months if the first child is a boy and is 45 months if the first child is a girl, the predicted duration between child 2 and child 3 is 59 months if the second child is a boy and 52 months if the second child is a girl and finally the predicted duration between child 3 and child 4 is 45.5 months if the third child is a boy and 43.6 months if the third child is a girl. In each of these cases the predicted duration was computed for an Indian woman, living in Natal and belonging to the lowest income class with the other explanatory variables held constant at their sample means. The estimated hazard functions (for an Indian woman, living in Natal and belonging to the lowest income class with the other explanatory variables were held constant at their sample means) are presented in Figures 1, 2 and 3 for the three Transitions. Note that at each transition the estimated hazard of

having another child is higher if the previous child is a girl. The results for the Indian households can therefore be summarised as follows [see also Table 6]:

1. The coefficient of *SEXPREV* is positive and significant at every transition. So irrespective of the number of children, Indian women delay having another child if their first born is a son.
2. In Transition 2  $\rightarrow$  3 the coefficient of *TOTMAL* is positive and significant implying that an increase in the total number of sons the woman has, significantly increases the duration between child 2 and child 3.
3. The different sex dummy *DIFFSEX* is not significant in any of the transitions.

While *DIFFSEX* is not significant for the Indian households, in the case of Transition 2  $\rightarrow$  3 the coefficient of *DIFFSEX* is significantly different from zero for the Coloured and White households. The coefficient of *DIFFSEX* is positive and significant implying that if child 1 and child 2 are of different sex, the duration between child 2 and child 3 is higher. In the case of Transition 3  $\rightarrow$  4 *DIFFSEX* is positive and significant only for the White households. [See Tables 5a, 5b and 5c].

Our results appear to refute the usual arguments for son preference - contributing support for elderly parents and contributing to household income and instead suggest the importance of religious beliefs and social customs (dowry system, lineage, familial and kinship ties etc.) in directing parental preferences. As argued earlier it is not particularly surprising that the Indian households behave differently from the others in South Africa. The male dominated (paternalistic) *kutum* which forms an extremely important aspect of Indian life in South Africa possibly plays an important role in the preference for the male child. Additionally, the traditional Indian custom

of dowry (where the family of the bride pays the family of the groom) as opposed to the custom of bride price in Black and Coloured households (where the groom pays the so-called bride-price to the family of the bride) could also be a major reason for such different parental preferences for the Indian households. Conversations with economists at SALDRU seem to corroborate this view.<sup>11</sup>

## 6 Robustness

We test for robustness by estimating the model using alternative explanatory variables. The inclusion of income category dummies as explanatory variables in the previous section enabled us to differentiate the behaviour of women across the different income categories. However, it could be the case that the behaviour of individuals within each category is also different, depending on the number of existing sons, sex of the previous child and the sex composition of the existing children. To account for this, we interact the income dummies with each of the gender preference variables as follows:

$$ISEXPINC1 = INC1 \otimes SEXPREV$$

$$ISEXPINC2 = INC2 \otimes SEXPREV$$

$$ITOTMINC1 = INC1 \otimes TOTMAL$$

$$ITOTMINC2 = INC2 \otimes TOTMAL$$

$$IDIFFSINC1 = INC1 \otimes DIFFSEX$$

$$IDIFFSINC2 = INC2 \otimes DIFFSEX$$

---

<sup>11</sup>Private conversations with Dudley Horner and Francis Wilson at SALDRU, University of Capetown. Further given that more than 60% of all Indians in South Africa in 1993 were Hindus, this argument seems to make sense.

The results for the Indian households are presented in Table 7. The interaction variables have no significant impact on the duration between the first and the second child. In the case of Transition  $2 \rightarrow 3$ , *ITOTMINC2* is positive and significant - this implies that for women belonging to the second income category, an increase in the number of existing sons increases the duration between the second and third child. However in this case *SEXPREV* and *TOTMAL* do not significantly affect the duration between child 2 and child 3. The effects of the interaction variables are not any stronger in the case of Transition  $3 \rightarrow 4$ . Here *IDIFFSINC2* is negative and significant. However in this case *SEXPREV* is not significant, but *TOTMAL* and *DIFFSEX* are. Therefore we find that the greater the number of sons the lower is the duration between child 3 and child 4 and mothers who have children of different sexes delay conceiving for the fourth time. There appears to be a switch in preferences - Indian women after reaching a target number of sons, now want to have a daughter.

It could be argued that in estimating duration between births, the use of household income as a measure of household quality endowment could lead to endogeneity problems - decisions regarding duration between births might affect labour supply decisions of women and that in turn might affect household income [see Angrist & Evans (1998)]. To avoid this potential endogeneity problem we consider alternative specifications for the household quality endowment variable. The two alternatives that we consider are the highest level of education attained by the mother and the highest level of education attained by the household head. Note that even the use of the highest level of education attained by the mother could lead to similar endogeneity problems.<sup>12</sup> The highest level of education attained by the household head is measured by three dummy variables *PRIMSCH*, *PRIMPLUS* and *SECONDAR* (*PRIMSCH* = 1 if the highest level of

---

<sup>12</sup>Actually a complete model would require a joint estimation of education level attained, fertility decisions and labour supply decisions. For the purposes of this paper we do not conduct the joint estimation.

education attained by the household head is primary school and 0 otherwise;  $PRIMPLUS = 1$  if the highest level of education attained by the household head is more than primary school but not secondary school and 0 otherwise;  $SECONDAR = 1$  if the highest level of education attained by the household head is secondary school or above and 0 otherwise).<sup>13</sup> Likewise the highest level of education attained by the mother is measured by three dummy variables  $EDUCM1$ ,  $EDUCM2$  and  $EDUCM3$  ( $EDUCM1 = 1$  if the highest level of education attained by the mother is primary school and 0 otherwise;  $EDUCM2 = 1$  if the highest level of education attained by the mother is more than primary school but not secondary school and 0 otherwise;  $EDUCM3 = 1$  if the highest level of education attained by the mother is secondary school or above and 0 otherwise).<sup>14</sup> The results are robust with respect to the choice of the measure of household quality endowment that we use. In particular in the case of the Indian households the gender preference variables are exactly of the same sign as in the bench mark case (where we use household income dummies as a measure of the household quality endowment).

It can also be argued that employment opportunities for women would influence the duration between births, with women who are employed having longer duration than women who are not employed. The data we have on employment is not retrospective in nature, i.e., the question asked in the survey addresses whether the mother is currently employed, not when the child was born. We include  $EMPLOY$  ( $EMPLOY = 1$  if the woman is currently employed, 0 otherwise) in our robustness check to examine if this variable helps in explaining duration between children. We find that  $EMPLOY$  is not significant, a result which is not very surprising given the nature of the data available on employment. Moreover the inclusion of  $EMPLOY$  does not have any effect on gender preference. Another variable that could affect duration between children is

---

<sup>13</sup>The reference category is that the household head has no education  $NOEDUC$ .

<sup>14</sup>Once again the reference category is that the mother has no education  $EDUM0$ .

whether the woman lives in an extended (joint) family or a nuclear family. In an extended family there would be more members who can take care of the children, hence the mother is not the only care giver. We define a variable called *EXTEND*, ( $EXTEND = 1$  when the woman lives in an extended family, 0 otherwise). We also include the total number of women in the household (*N\_FEM*) as a measure of the “extra help” available to the mother. Our results indicate that an increase in the total number of women in the household reduces the duration between children (except in the Transition  $3 \rightarrow 4$ ), implying that women in households with more women wait less to have another child. *EXTEND*, however comes out to be positive and significant, hence women in extended families wait more to have another child. These two results seem to be contradicting each other, however this could be explained by the fact that extended families could already have a large number of children in the household and women might utilize this information while deciding on the duration between births. To check this we include the number of children in the household (*N\_CHILD*) instead of the number of women and define an interaction term ( $INCEXT = N\_CHILD \otimes EXTEND$ ) between *EXTEND* and number of children (*N\_CHILD*). Having more children in the household increases the spacing between child one and two and the interaction term is also positive and significant, hence women in extended households with higher number of children have longer duration between births. This interaction term is not significant for the other two transitions. The important point to note here is that irrespective of the choice of the explanatory variables used, the gender preference variables are always significant and positive in explaining duration between children. The results regarding son preference in Indian households are therefore fairly robust.



## 7 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, South Africa has never been examined. The presence of the four distinct races in the country makes it an interesting country to study and the presence of a strong Asian (Indian) community within the country sets the stage for an interesting comparison with the Asian countries. We use 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 child. For the non-Indian households we do not find any evidence of son preference (at any stage of life).

The implications of our results are very interesting and in some ways a bit disappointing as it seems that economic development in itself is not enough to overcome gender preferences. Usual economic arguments to explain son preference include the old age security motive for having sons or the fact that in rural households sons can provide an additional hand in the farms. One expects that as the economy develops, the availability of alternative means of old age pension would reduce the old age security motive for preferring sons and the availability of improved farming technology would reduce the demand for child labour on farms. In South Africa, no Indian household resides in rural communities and there is a means-tested social pension program that reduces the dependence on children for old age support. Hence there appears to be little economic rationale for preferring sons over daughters. One has to therefore acknowledge the role

of customs and traditions in directing parental preferences regarding the gender composition of their children.

## 8 References

1. Angrist, J. D. & Evans, W. E. (1998), 'Children and their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size', *American Economic Review*, **88**(3), 450-477.
2. Ali Khan, M. & Sirageldin, I. (1977), 'Son Preference and the Demand for Additional Children in Pakistan', *Demography*, **14**(4), 481 - 495.
3. Aly, H. S. & Shields, M. P. (1991), 'Son Preference and Contraception in Egypt', *Economic Development and Cultural Change*, 353 - 370.
4. Arnold, F., Choe, M. K. & Roy, T. K. (1998), 'Son Preference, the Family Building Process and Child Mortality in India', *Population Studies*, **52**, 301 - 315.
5. Chowdhury, M. K. & Bairagi, R. (1990), 'Son Preference and Fertility in Bangladesh', *Population and Development Review*, **16**(4), 749 - 757.
6. Das, N. (1987), 'Sex Preference and Fertility Behavior: A Study of Recent Indian Data', *Demography*, **24**(4), 517 - 530.
7. Gangadharan, L. & Maitra, P. (1999), 'Two Aspects of Fertility Behaviour in South Africa', *University of Melbourne Working Paper # 683*.
8. Haughton, J. & Haughton, D., (1998), 'Are Simple Tests of Son Preference Useful? An Evaluation using Data from Vietnam', *Journal of Population Economics*, **11**, 495 - 516.
9. Larsen, U., Chung, W. & Das Gupta, M. (1998), 'Fertility and Son Preference in Korea', *Population Studies*, **52**, 317 - 325.
10. Lemon, A. (1987), *Apartheid in Transition*, Gower Publishing Company, Vermont, USA.

11. Leung, S. F. (1988), 'On Tests for Sex Preferences', *Journal of Population Economics*, **1**, 95 - 114.
12. Rahman, M. & Da Vanzo, J. (1993), 'Gender Preference and Birth Spacing in Matlab, Bangladesh', *Demography*, **30**(3), 315 - 332.
13. 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.
14. Rosenzweig, M. R. (1986), 'Birth Spacing and Sibling Inequality: Asymmetric Information Within the Family', *International Economic Review*, **27**(1), 55 - 76.
15. SALDRU, Project for Living Standards and Development, (1994), *South Africans Rich and Poor: Baseline Economic Statistics*, School of Economics, University of Capetown.