Marriage Trends in South Africa: 1995-2006

Grace Kumchulesi
University of Malawi, Chancellor College

September 16, 2010

Abstract

Marriage rates for African women aged between 15 and 49 years declined from 35 percent in 1995 to 31 percent in 2006. Using datasets from nationally representative surveys from 1995 to 2006, we constructed a synthetic panel where various birth cohorts of women were followed through the years.

We employ the Age-Period-Cohort model to establish if this marriage decline is not driven by erratic fluctuations in sampling designs of independent surveys.

The estimates of age effect show that the odds of marriage for women increase with age, just like that of their predecessors. Cohort effects are also positive, indicating that the older the birth cohort is, the higher is the likelihood of marriage. Period effects are smaller in magnitude. In addition, the year on year differential shows no peculiar behaviour in any of the survey years. This suggests that there were no outlier surveys in sampling of African women.

1 Introduction

More research on the decline in marriages has been undertaken for developed countries than for developing countries. In the United States of America for example, most of these studies compare black women with white women. The findings show that while the decline in marriages is pervasive throughout the US society, the extent is greater in the African American population group (Tucker and Mitchell-Kernan, 1995).

In South Africa, similar evidence has been documented. Marriage rates from datasets of the earlier surveys are higher than those from the more recent surveys. Most researchers have interpreted this as a decline in marriage. However, since these surveys are independent and sampling designs change, it is possible that the observed marriage decline is a mere sampling issue, where possibly, some surveys sampled more married women than others. This study makes use of repeated cross-sectional survey data, spanning over a decade. Using this dataset, this paper will undertake a trend analysis, in order to establish if the marriage decline is real.

At the exploratory stage of the analysis, we compare the marriage rates from the repeated cross-sectional surveys. This is done as a preliminary check on whether marriage rates appear to have changed between 1995 and 2006. A bivariate relationship between age and marriage is also evaluated at this stage. Further, we "pool" the cross-sectional surveys in a cohort analysis framework in order to determine whether the relationship between age and marriage has changed over time. This undertaking demonstrates whether the observed marriage decline in the cross-sectional analysis may be a generational change, suggesting possible permanent change in marital behaviour among African women in the period under study. Finally, the paper undertakes a decomposition analysis for trends. This procedure aims at distinguishing trends from erratic fluctuations. At that stage, we decompose the marriage trends into age, period and cohort effects in order to determine whether the change in marital pattern is influenced by these.

The rest of the paper proceeds as follows. Section 2 briefly summarizes prior evidence of decline in marriages in South Africa. In section 3, we present a detailed discussion of the national surveys to be used in the analyses. We also highlight some of the concerns associated with use of such data. Section 4 presents the empirical results from the descriptive analyses of marriage trends in South Africa from 1995 to 2006. In section 5, we discuss the methodology used to disentangle the age, period and cohort effects, followed by a presentation of the results from the decomposition analysis. Finally, section 6 concludes the paper.

2 Prior Evidence of Declining Marriages in South Africa

Several researchers have documented declining marriages in South Africa. However, most of the research on this subject has shortfalls. For example, Casale and Posel (2002) and Casale (2003) picked up the decline while studying feminisation of the South African labour market. Since marriage was not the core of their studies, the marriage rates that are reported are too aggregated.

Others like Kalule-Sabiti, Palamuleni, Makiwane and Amoateng (2007) focused on marriages. Using data from Census 1996, Census 2001 and South Africa Demographic and Health Survey 1998, Kalule-Sabiti et al. (2007) found that marriage is more popular among whites and Asians than among African and coloured population groups. Marriage also seems to be unpopular among younger women and more educated women. However, they used data from different sources to generate marriage trends, and the marriage decline from such data is likely to suffer from change in sampling designs.

Going back to the pre-democratic South Africa, a number of ethnographic studies also noted a “crisis” in the African family. Denis and Ntsimane (2004) and Hunter (2005a), for example reveal extremely low marital rates among Africans, especially from the 1980s. Aside from difficulties in interpretation (because African marriages are a process rather than an event), apartheid statistics are notoriously unreliable because of under-representation of Africans in surveys.

A more recent study by Hosegood, McGrath and Moultrie (2009) found that marriages have continued to decline even in the 2000s. However, this study had a narrowed scope as it examined marriage trends only in the KwaZulu Natal Province of South Africa, which may not be representative of South Africa.

Difficulties with these studies mean this study can make a significant contribution to the literature on marriages in South Africa. This study uses datasets from a series of repeated cross-sectional surveys from 1995 to 2006, which makes our series longer and smoother. Secondly, the surveys from which the samples are drawn are nationally representative, and finally, the surveys were conducted by one data producer, Statistics South Africa, which rules out biases arising when data is collected by different organisations.

For these reasons, our study may be viewed as an update of marriage trends in South Africa. Most importantly, is the decline in marriages even real? In other words, is the marriage decline just a sampling issue, where possibly and for whatever reasons, some surveys sampled more married women than others? Our use of a series of independent cross-sections and comprehensive decomposition analysis on marriage rates make this study important. In the following section, we discuss in detail the data to be used as well as its limitations.

3 Survey Data

South Africa does not have a database that captures marriage histories in detail while at the same time being nationally representative. The only survey that collects at least additional information on marriage data (other than marital status) is the 1998 KwaZulu Natal Income Dynamics Study (KIDS). KIDS resurveyed 1100 households from the KwaZulu Natal Province who were initially interviewed for the first nationally representative survey, the Project for Statistics on Living Standards and Development (PSLSD) in 1993. The main limitation of using the KIDS dataset for our purposes is that it is not nationally representative, because the sample is representative of KwaZulu Natal Province only. For this reason, we did not make use of the KIDS dataset in this study.

The other available datasets that are nationally representative are from the Census 1996 and Census 2001. However, the limitation of using census data for our purposes is that the South African census is only conducted five-yearly. For that reason, we remain with data collected only at two points in time. This would challenge the basic analytic principle that data cumulated at more time periods is more useful for the analysis of trends. A firmer basis for conclusion regarding trends would therefore be provided by many repeated surveys cumulating a decade or more of annual or biannual data. This study has been made possible by availability of such data in South Africa.

Our main sources of data are the 1995-1999 October Household Surveys (OHSs) and the September wave of the 2000-2006 Labour Force Surveys (LFSs). Although this series of surveys was initiated in 1993, we have left out the 1993 and 1994 datasets because their sampling methodology significantly differs from the 1995 survey onwards. In 1993 and 1994, a sample of 30000 households was drawn from 1000 enumeration areas. From 1995 onwards, the sample was drawn from 3000 enumeration areas. Moreover, in 1993, the TBVC states[1] were not included in the sample. Varying sampling methodologies like these challenge survey to survey comparisons. For these reasons, we work with the more consistent cross-sectional

[1]Transkei, Bophuthatswana, Venda and Ciskei were bantustans or black African homelands.

dataset from 1995 onwards.

The OHSs are annual independent cross-sectional surveys, and different samples were designed for each of them. However, in a standard OHS, the sample was explicitly stratified by province, magisterial district, urban/rural and population group. A sample was drawn by applying a two-stage sampling procedure. In the first stage, enumeration areas (EAs) represented primary sampling units (PSUs) were systematically selected by means of probability proportional to size principles in each stratum to ensure adequate representation. The measure of size was the number of households in each PSU. The database of EAs, as established during the demarcation phase of Census 1996, constituted the sampling frame for selecting EAs. In the second stage, households were randomly selected from the selected EAs. Depending on the availability of funding, the number of sampled households ranged from 16000 in 1996 through to 20000 in 1998 and 30000 in 1995, 1997 and 1999.

The LFS replaced the OHS after their discontinuation in 2000. The sampling design for the standard LFS is similar to that of the OHS. The main feature distinguishing these two surveys is that the LFS were designed to be a twice yearly rotating panel with the waves running in March and September. A twenty percent rotating scheme was designed, implying that new dwelling units would be included to replace the dropped ones in the second wave (SADA, 2001). For the purpose of this study, we will only make use of cross-sectional information and the September wave in particular.

The preceding paragraphs immediately suggest that comparisons both between and within the two surveys might not be simple. Data from both surveys were collected by Statistics South Africa, which rules out biases resulting from "house effects" (Johnston, 1981), especially when detecting a trend from survey to survey. Data from a single survey organisation is not immune from error, however. Even if data is collected by the same organisation, the wording of a question might have changed from survey to survey, producing a significant change in response (Rasinski, 1988; Smith, 1987). "Context effect" is another concern, in which responses to a particular question may be affected by the nature of that question as well as that of other questions in the survey (Smith, 1988; Smith, 1992). Other than changes in sampling methodology and consistency issues, another concern may be changes in the definition of variables over time.

One major limitation of the OHSs and LFSs is little coverage of questions on marriage, in particular, and other questions generally useful in a marriage study. For instance, marriage dates are missing. In addition, a highly culturally diverse society like South Africa would be expected to have culturally influenced marital behaviors. Unfortunately, these surveys did not ask questions which would accommodate a culturally disaggregated study. For example, brideprice (or *ilobolo* payment) is one of the important features that differentiate nuptials between different ethnicities. At best, one could examine racial differentials in marriage rates, but this would not be adequate to establish marital behavior of a particular ethnic group, for example. Nonetheless, despite these and other data limitations, the study proceeds to give an aggregated picture of marriage trends, disaggregated by age.

For trend analysis, we use synthetic panel data constructed from twelve successive cross-sectional national surveys. For each year of the survey data, our variable of interest is responses to the question on marital status. The OHS has six categories namely: married (civil), married (traditional/customary), living together with partner, widower/widow, divorced/separated and never married and the categories represented in the LFS include married or living together as husband and wife, widow/widower, divorced/separated and never married. StatsSA (2001) defines marriage to include marriages by civil or religious ceremonies, marriages by customary law, and all those living together as married. In this regard, and for consistency with the classifications in the LFS, we will merge the first three categories in the OHS as married. We will therefore estimate the proportion of a particular age cohort for each marital state.

We restrict our sample to African women between the ages 15 and 49, since this is the prime age range in which marriages take place. Nevertheless, we also present marriage rates for white women and African men for the purpose of racial and gender comparisons.

4 Marriage Patterns: The Relationship between Age and Marriage

We commence with a descriptive analysis where we present the cross-sectional and cohort dimensions of marriage patterns, while uncovering the relationship between age and marriage. To estimate the proportion of a particular age or age cohort that falls into one of the four marital categories, we employ locally weighted estimations of a scatter-plot smoothing (LOWESS) technique, which is explained in detail in appendix A. This technique was developed by Cleveland (1979) and is used to smooth scatter-plots which have been contaminated with noise. The advantage of using nonparametric estimation is that it does not impose a particular functional form on the data for relationships but allows the data to reveal the form of a relationship (Deaton, 1997). Thus, no prior assumption about a relationship is made.

4.1 Stylized Facts from Cross-sectional Analysis

We first present findings from the cross-sectional analysis, giving visual evidence of declining marriage rates in the period 1995 to 2006. We compare marriage trends across gender and racial lines, while focusing on African women. From Table 1 in appendix C, we observe that while marriages were generally declining across all races, the decline is astounding among Africans. In 1995, the marriage rate was 35 percent for African women aged between 15 and 49 years old and by 2006, this had dropped by about 10 percent change to 31 percent.

The relationship between age (in years) and marriage for African women, overtime, is depicted in the top panel of Figure 1. Each trajectory in the figure (drawn using LOWESS technique) depicts the proportion of married African women for a particular age in a particular year[2]. We observe clear evidence of declining marriage rates over the period 1995-2006, across age.

The balance at each age for each year for the married state is single, widowed or divorced, implying that a decline in the married state is an increase in either or all of the other marital states. The diagram for the single state in the bottom panel of Figure 1 nearly mirrors the one for married state, for all ages, with the deviations captured in the widowed and divorced categories. This suggests that non-marriage is what adequately accounts for the declining marriage rates, as opposed to singlehood resulting from marriage dissolution by death of a spouse or divorce. The cross-sectional picture provides a snapshot pointing towards a decline in marriages. The decline is evident from age 25 and beyond.

For each cross-section, the proportion of married women increases with age, and rapidly so, until the age of 35 years, after which it remains fairly stable. This suggests that when African women get to the mid-thirties, they are unlikely to enter into marriage. The surveys did not ask questions about age at marriage, making it difficult to establish whether there is any evidence of late entry into marriage. However we find that for each cross-section, a significant proportion of African women start getting married after the age of 22 years, giving evidence of late entry into marriage. Racial calculations of the singulate mean age at marriage[3] in Table 2 (in appendix C) shows that African women enter marriage at relatively older ages compared to their white counterparts. In addition, the singulate mean age at marriage for African women is at least 29 years in all the years, indicating evidence of late marriages. This conclusion is based on a proposition made by Bogue (1969)[4], that the median age of 22 years marks the benchmark for late entry into marriage for females. Gender comparisons of the singulate mean age at marriage indicates that the married male sample is on average older that the married female sample across all races and years,. This is unsurprising, considering that a man is usually older in a marriage.

One of the limitations of this study is failure to separate out the extent of cohabitation in the married state due to inconsistencies in the categorization of the marital states for the different surveys. For instance, out of the twelve surveys, married and cohabitating states were lumped together from the years 2000 to 2003. In that regard, marriage rates would be inflated by cohabitation, and the decline in marriages may be understated. This kind of data capturing thus becomes a limitation if one wants to investigate whether marriages are being replaced by single-hood or by cohabitation.

We show, in Table 3 in appendix C, the extent of this problem, by factoring out cohabitation from the married state in the surveys where the two were categorized separately.

[2]In tracking marital states across time, we noticed an inconsistency in the 1998 OHS. Marital status codes were inadvertently swopped. The code for "married" was used for "never married"; "never married" for "divorced"; "divorced" for "widowed"; and "widowed" for "married". The inconsistency was a result of incorrect documentation that had been sent out with the survey. The 1997 questionnaire had been distributed with a 1998 cover sheet. However, the marital status codes had changed in the 1998 questionnaire and this led to erroneous coding of the categories. Errors like these are critical, and especially because marital status is one variable that is usually included in most regression analyses.

[3]For pedagogy, step by step calculation of the singulate mean age at marriage is shown in appendix B.

[4]In Kalule-Sabiti et al. (2007).

Married proportions are clearly inflated when cohabiters and the married are lumped together. Cohabitation over the twelve-year period has increased by 7 percentage points. This suggests that African women's union formation behaviour may be similar to that of the western world, with informal unions replacing the formal unions. However, whether cohabiting relationships in South Africa tend to be long-term or short-term is another issue which needs further investigation.

The evidence from the cross-sectional analysis suggests that marriages have declined, not only among African women, but also across gender and racial lines. We further test the hypothesis that the decline in marriages of African women is real by using cohort data,

4.2 Stylized Facts from Cohort Analysis

In cross-sectional analysis, we observe an individual only at a point in time. However, individuals are aging throughout the cross-sections with possibility of cohort replacement and different marital behaviours for different cohorts. Such changes cannot be picked up from cross-sectional analysis. With a rich series of cross-sections, however, we are able to follow cohorts of individuals throughout the cross-sections. This helps to understand whether the changes observed in marriage rates are generational changes. Cohort analysis, involving comparing an age cohort in different cross-sections is more informative than simply comparing subjects of the same age in different years.

We define cohort by age in 1995, and we follow these age cohorts through the eleven years to 2006. The picture from the selected cohorts in Figure 2 shows that within cohorts there is a positive relationship between age and proportion of women that are married. Between cohorts, we observe that for most part of the period, trajectories of older age cohorts lie above the ones for the younger cohorts, suggesting different marital experiences for different generations. For instance, comparing the 20 year old age cohort at the age of 25 and the 24 year old age cohort at the same age, we observe a clear shift in marital behaviour, with the proportion married for the younger cohort, lower than their predecessors. Similar comparisons can be made between other cohorts, with results pointing to a decline in marriages.

The trajectories are not smooth, demonstrating cohort-global effects. With survey data, concern is more around sampling effects rather than macroeconomic effects. By grouping cohorts in five-year bands, to smooth the trajectories, shown in the bottom panel of Figure 2, we obtain clear evidence of differences in marital behaviour between the younger and older women. At the age of 35, 50 percent of the cohort aged 20-24 in 1995 is married, while at the same age, about 65 percent of the cohort aged 30-34 in 1995 is married. We therefore conclude from the cohort analysis that within cohorts, the proportion married increases with age, and between cohorts, marital experiences are different, signifying a generational change. However, we do not rule out the presence of sampling effects, following the zigzagged trajectories in the top panel of Figure 2.

The picture from the cohort analysis includes age, cohort and period effects. The age and cohort effects are clearly seen. With regard to the cohort effects for example, the trajectories for the older cohorts are [always] above the younger cohorts (in the five-year age band except for the 15-19 and 20-24 age groups, which coincide). This suggests that the young and the old face different marital experiences, signifying a generational change.

The age effect is also evident by the positive relationship between age and proportion married and takes a slight S shape. In this study, it is evident that the proportion of women who are married tends to increase rapidly between ages 25 and 30, and less rapidly beyond 30. Though the age effect is discontinued for the later ages in this study, the proportion of married women is expected to be fairly constant in the ages beyond 50 when no or less marriage formations take place, already formed marriages become stable and deaths stabilize too. The proportion is expected to start falling in old ages when marriage dissolutions, usually by death, have little or no replacement by new entrants into the marriage institution.

The period effects are captured by the span of each trajectory, which corresponds to the 12 years of the surveys, 1995 to 2006. Graphs from the Age-Period-Cohort models disentangle age, period and cohort effects from marriage rates.

5 Decomposition Analysis: Age, Period and Cohort Model

An event such as marriage is influenced by age, period and cohort characteristics, in some way. If we observe that marriage rates are declining in datasets from independent surveys, the decrease could be due to differing sampling designs (period effect), more specific processes related to aging (age effect) or if specific cohorts display differing likelihoods to marry, then the decrease could be cohort-based. A more revealing picture of underlying marriage rate schedules can therefore be extracted by disentangling age, period and cohort effects.

The Age-Period-Cohort (APC) model[5] is the generational model used to identify whether changes in human behaviour are cohort-based or due to other factors such as age or calendar year. In marriage analysis, age effect would represent different allocations between the marital states associated with different age groups, giving a typical age profile. The period effect represents variation in distribution across marital states over time, simultaneously associated with all age groups, and affects every cohort. Cohort effect represents the difference in the distribution across groups of individuals with the same birth year (Fu, Hall and Rohan, 2004).

However, attempts to separate age, period and cohort effects are bedevilled by the identification problem resulting from specifying all three in an additive model[6]. Due to the exact

[5]In addition to empirical application on data from repeated cross-section surveys, APC models are also applicable to analysis of age-by-time period tables of rates (for example Fu, 2000; Yang, Fu and Land, 2004; Yang, Schulhofer-Wohl, Fu and Land, 2008; Yang, 2008) and in cohort analysis of accelerated longitudinal panel designs (for example Miyazaki and Raudenbush, 2000; Yang, 2007).

[6]$Cohort = Period - Age$. However in an age-period-cohort regression model, neither of the age or period or cohort effect can be estimated, *ceteris paribus*. For example, if cohort is defined by year of birth and κ denotes the expected change in marriage rate when year of birth is increased one year, κ cannot be estimated

linear dependency nature of the APC model, a unique solution for each of the effects cannot be found, since matrix containing age, period and cohort variables does not have full rank.

There is no perfect technical solution to disentangle age, period and cohort effects. However, a number of defensible solutions to this problem have been proposed in the literature (for example Glenn, 1976; Fienberg and Mason, 1978; Hobcraft, Menken and Preston, 1982; Wilmoth, 1990; Deaton, 1997; O'Brien, 2000; Fu, 2000; Knight and Fu, 2000; Yang et al., 2004; Yang, 2006; Yang and Land, 2006), one of the simplest solutions being to drop one of the multicollinear variables. This strategy assumes that one of the effects is zero, which however, becomes problematic where all of the variables are seen to be potentially important in explaining behaviour. This means the results are only as reliable as the identification assumption made, and for example, if age effect is wrongly assumed to be zero, the estimates of period and cohort effects will be contaminated with age effects. That poses problems for correctly interpreting the two biasedly estimated coefficients.

A widely used class of correctional methods that acknowledges the importance of all three involves making constraints on the coefficients, generally known as the coefficient-constraint approach[7], where constraints suiting the phenomenon under study have to be imposed. As with dropping one of the effects, the problem with this methodological remedy is that the parameter restrictions have to be theoretically motivated, which poses a challenge for an analysis of most phenomena including the current study. In addition, the estimates are sensitive to the choice of restrictions and goodness-of-fit to the data of constrained models is impossible to determine.

A number of authors have proposed different identifying assumptions. Deaton (1997) suggests a normalisation which assumes that period effects have zero mean and are orthogonal to the time trend. For example, here, where marriage proportion is the variable to be decomposed, the idea is to attribute marriage rate differentials between the surveys to age and cohort effects, and to use period effects to capture effects from changes in sampling designs that averages to zero in the long run. Empirical estimation[8] of the APC model subject to Deaton's (1997) normalisation involves the usual dropping of a row each of the age and cohort dummy matrices. In addition, two rows of the period dummy matrix are dropped so that the remaining $T-2$ period dummy matrix is defined according to:

$$d_t^* = d_t - [(t-1)\, d_2 - (t-2)\, d_1] \tag{1}$$

where d is a dummy variable taking the value of 1 if period is t, and zero otherwise, for $t = 3, ..., T$. The period effects for the two previously dropped period dummy variables are recoverable from the fact that all period effects sum to zero and are orthogonal to the time trend.

A relatively new approach to the estimation of APC models, called the intrinsic estimator,

---

for a fixed age and period. Likewise, age slope (or period slope) can not be estimated for a fixed period and cohort (fixed age and cohort). This is what is refered to as the identification problem in cohort analysis.

[7]Firebaugh (1997) calls this "technical fix" strategy or "fixing coefficients *a priori*".

[8]For a stata procedure of Deaton's normalization please refer to Deaton (1997, pp. 407).

has been developed. The intrinsic estimator approach utilizes estimable functions that are invariant to the selection of constraints on parameters[9] (Fu, 2000). This approach has been validated by both case studies and simulation tests of model validation[10] (for example Yang et al., 2004).

The second strategy is the proxy variables approach. This strategy uses one or more variables to proxy the age, period, or cohort coefficients (for example O'Brien, 2000). Difficulty with this approach is that it may not be in the analyst's interest to assume that all of the variation associated with the age, period and cohort dimensions is fully accounted for by a proxy variable.

The third approach is the non-linear parametric (algebraic) transformation approach, where a non-linear parametric function of one of the age, period, or cohort variables is defined so that its relationship to others is non-linear. A limitation with this strategy is determining what non-linear function to be defined for the age, period, or cohort effect.

Clearly, there is no "magic formula" for separating age, period and cohort effects. With no strong hypothesis to motivate the influence of age, period and cohort effect on marriages in South Africa, this study prefers Deaton's identifying assumption as a descriptive device for expressing marriage trends.

5.1 Model Specification and Estimation

In linear regression form, the APC model for marriage rates can be specified as:

$$M_{ij} = \mu + \alpha_i A + \gamma_j P + \kappa_k C + \epsilon_{ij} \tag{2}$$

where M_{ij} denotes mean marriage rates for the ith age for $i = 15, ..., 49$ at the jth time period for $j = 1995, ..., 2006$; μ denotes the intercept; α_i denotes the ith age effect; γ_j denotes the jth period effect; κ_k denotes the kth cohort effect for $k = 1, ..., 35$ cohorts, with $k = i + j - 1995$; and ϵ denotes the random error term with $E\left[\epsilon_{ij}\right] = 0$ and $E\left[\epsilon_{ij}^2\right] = \sigma^2$. A, P and C are matrices of age, period and cohort dummies, respectively.

Due to the grouped nature of the data, the dependent variable is a proportion, rendering estimation by ordinary least squares (OLS) inappropriate. A better alternative which we adopt for this study is to estimate a generalized linear model (GLM), a method proposed by Papke and Wooldridge (1996). This is analogous to using a logit or probit model in place of a linear probability model, in order to ensure that the predicted probabilities are in the valid range of zero and one. By estimating a GLM instead of an OLS, the predicted values of the fractional dependent variable will fall between the range of zero and one. Accordingly, we estimate age, period and cohort effects by quasi-likelihood estimation methods which do not need any special data adjustments for the extreme values of zero and one (Papke and Wooldridge, 1996). The quasi-likelihood method is fully robust and relatively efficient under

[9] Refer Yang et al. (2008), for the algebraic, geometric and verbal description of the intrinsic estimator as well as its application in stata.

[10] Including estimability, unbiasedness, efficiency and asymptotic consistency.

the GLM assumption. In this regard, we maximise a Bernoulli log-likelihood function, given by:

$$l_i(\beta) = M_{ij} \log\left[G(\mathbf{x}_i\boldsymbol{\beta})\right] + (1 - M_{ij}) \log\left[1 - G(\mathbf{x}_i\boldsymbol{\beta})\right] \tag{3}$$

where $G(.)$ is a logit link function, ensuring that predicted values of the dependent variable lie in $(0, 1)$; β is a vector of parameters of age, period and cohort effects, $\mathbf{x}_i's$ are matrices of age, period and cohort dummies; and M_{ij} denotes marriage rates as previously defined.

Taking the estimation issues into account, equation (2) takes a generalised linear model form with a logit link and a binomial family to get a logistic model formulated as:

$$\theta_{ij} = \log\left(\frac{m_{ij}}{1 - m_{ij}}\right) = \mu + \alpha_i A + \gamma_j P + \kappa_k C \tag{4}$$

where θ_{ij} is the odds of marriage and m_{ij} is the probability of marriage in cell (i, j). We include the robust option in the GLM estimation in order to obtain robust standard errors which is particularly useful if the distribution family is incorrectly specified. Subject to Deaton's (1997) normalization, the regression consists of $(I - 1)$ age dummies, $(J - 1)$ period dummies and $(K - 2)$ cohort dummies.

5.2 Age, Period and Cohort Effects

Following the Age-Period-Cohort decomposition strategy, we present, in Figures 3, 4 and 5 the estimated coefficients and their 95% confidence intervals plotted for successive categories within the age, period, and cohort classifications. But since no theoretically motivated restrictions were made, we use the APC model as a descriptive device, which is adequate for this paper's objectives.

The age effect shown in Figure 3 is positive for all cohorts and its magnitude increases with age. The global positive age effect indicates that the odds of marrying increase as young adults age, as with their predecessors. In other words, the age cohorts under study are merely adhering to what is expected of individuals as they age, in terms of marriage, which replicates the marital behaviour of prior generations. The weaker positive age effects for younger cohorts indicate that compared to older people, the odds for marriage for younger people are less. This is unsurprising, considering the life-cycle expectation of marriage. Most likely, the young adults are still in school, and having completed their schooling, they are more readily available for marriage. Consistent with the estimated mean age at marriage of 30 years, age effect shows that the age of 30 years is the peak-age for marriages. Beyond the age of the 30, age effect remains fairly constant or is weaker. Perhaps marriage behaviour after the age of 30 is due to women's economic independence, as they have by then established themselves in the workplace.

Figure 5 shows that cohort effects increase with age. The earlier a woman is born, the older she is in 1995, and age in 1995 defines the cohort an individual belongs to. We observe very low and near constant effects on the odds of marriage for women up to the age of 24, after which cohort effects increase. The implication is that cohort effects are present and the younger and older generations behave differently, driven by their own cohort influences. A potential explanation for the different marital behaviours is that the younger African women are enjoying the opportunities afforded to them by change in the political environment. Unlike the older generation, they are able and have economic incentives to stay longer in school, rather than marry at a young age as their predecessors did. In other words, while an event such as new government is common to all cohorts, its consequences regarding marriage vary between the younger and the older generations. The association between cohort and marriage is characterised by a decline in marriage rates in younger cohorts.

The estimated year effects in Figure 4 are much smaller in magnitude than either the cohort or age effects. In addition, the year on year differential of the period effect is negligible, with no peculiar behaviour in any of the survey years. Mostly yearly period effects are around zero. This result suggests that there are no outlier surveys as far as sampling of African women is concerned, and therefore, the cross-sections are comparable. This is a very important result because in confirms that the decline in marriages observed in the cross-sectional and cohort analyses above is a real incident which is not the result of sampling and questionnaire design. The decline in marriages possibly reflects a push for African women to achieve economic, educational and career goals, enhanced by the opportunities enabled by new legislation of the post-apartheid government supporting women's rights.

6 Conclusion

In this paper, we set out with information that marriages may be declining. Using data from a series of nationally representative household surveys conducted from 1995 to 2006, we endeavored to examine if the marriage decline deduced from comparing marriage rates at different time points of cross-sectional survey data is a real decline and not one driven by differing sampling designs used in the surveys. We constructed a synthetic panel from the October Household Surveys and the Labour Force Surveys. Cohorts, which we defined by year of birth (or age in 1995), were followed through the period.

Our main objective was to separate period effects from the age and cohort effects in the marriage rates in order to establish if marital experiences of younger and older cohorts are different. The period effects are negligible, suggesting that the marriage decline is a real decline. Age effects are present and greater in magnitude but, most importantly, cohort effects are also evident, indicating a shift in marital behaviour between the younger and the older cohorts, pointing to decline in marriage. We identify women's increased economic independence (afforded by more favourable labour market opportunities for African women owing to reforms in post-apartheid labour markets) and shortage of "marriageable" men (due to increased African male joblessness) as likely explanations for the decline.

In this paper, we have shown that that marriages are indeed declining. Marriage rates for African women aged between 15 and 49 years old declined from 35 percent in 1995 to 31 percent in 2006 and the period effect component of these marriage trends confirms that this decline is a real decline.

Figure 1: Married and Single African Women: 1995-2006

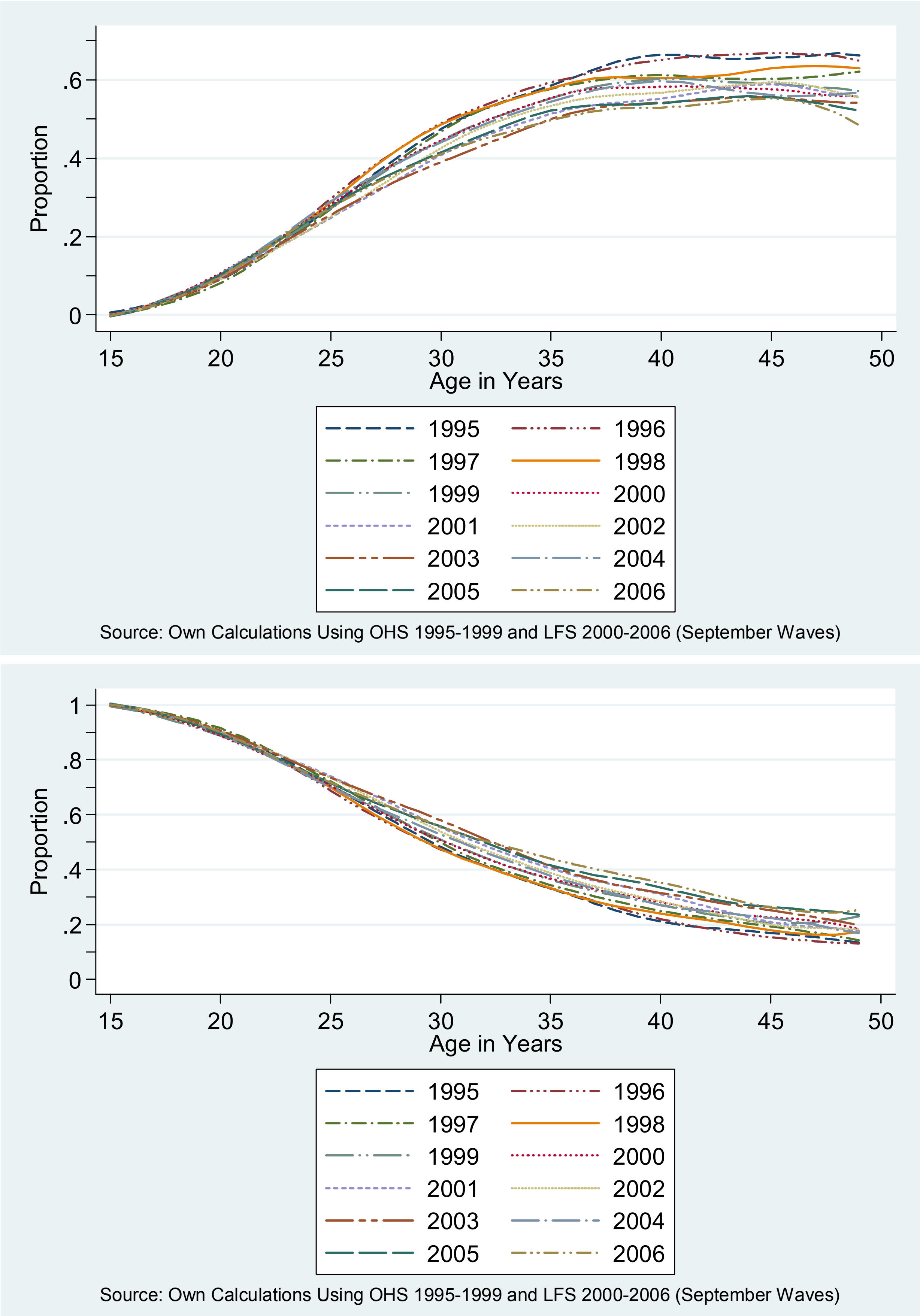


Figure 2: Proportion Married by Cohort

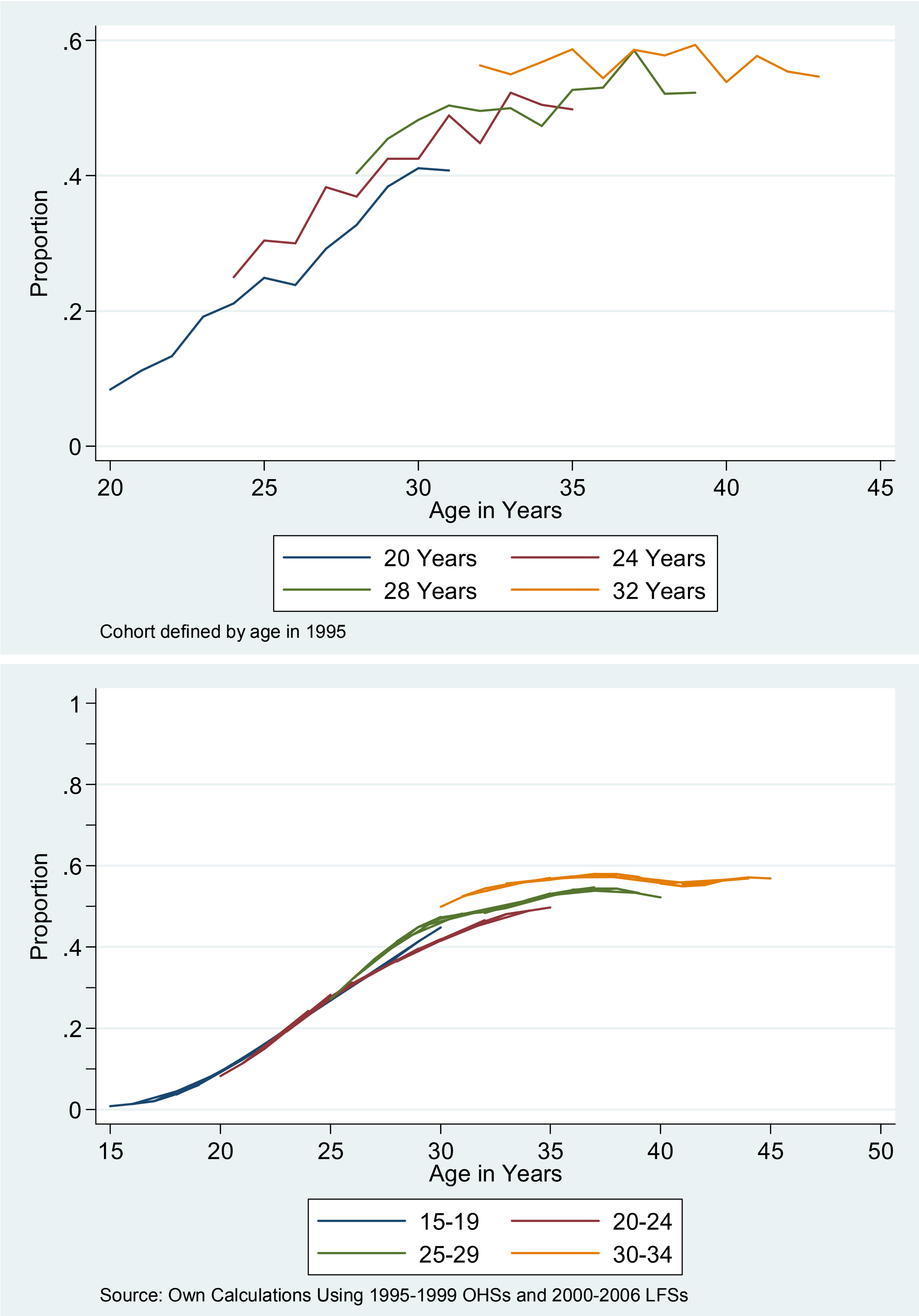


Figure 3: Coefficient Estimates and 95 Percent Confidence Intervals of the Age Effects on Marriage Trends

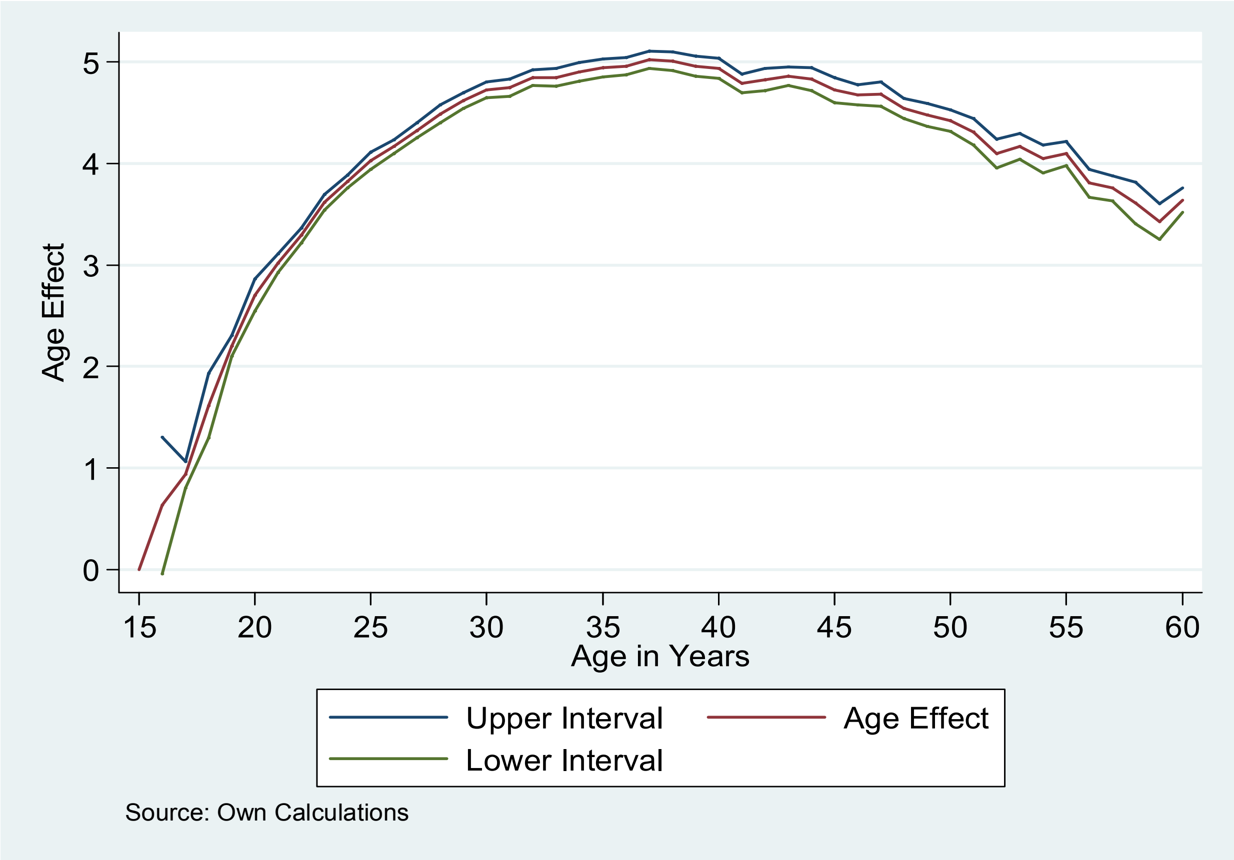


Figure 4: Coefficient Estimates and 95 Percent Confidence Intervals of the Period Effects on Marriage Trends

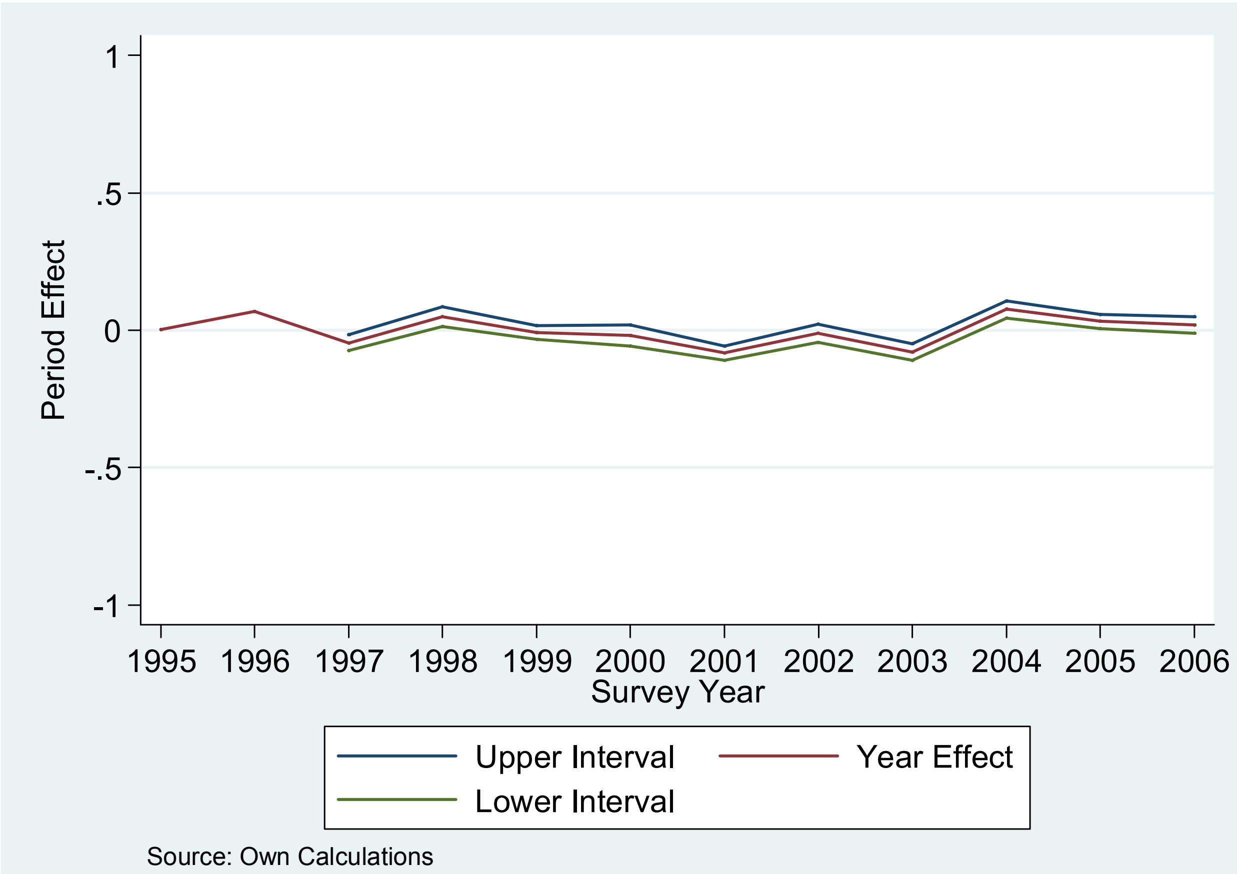
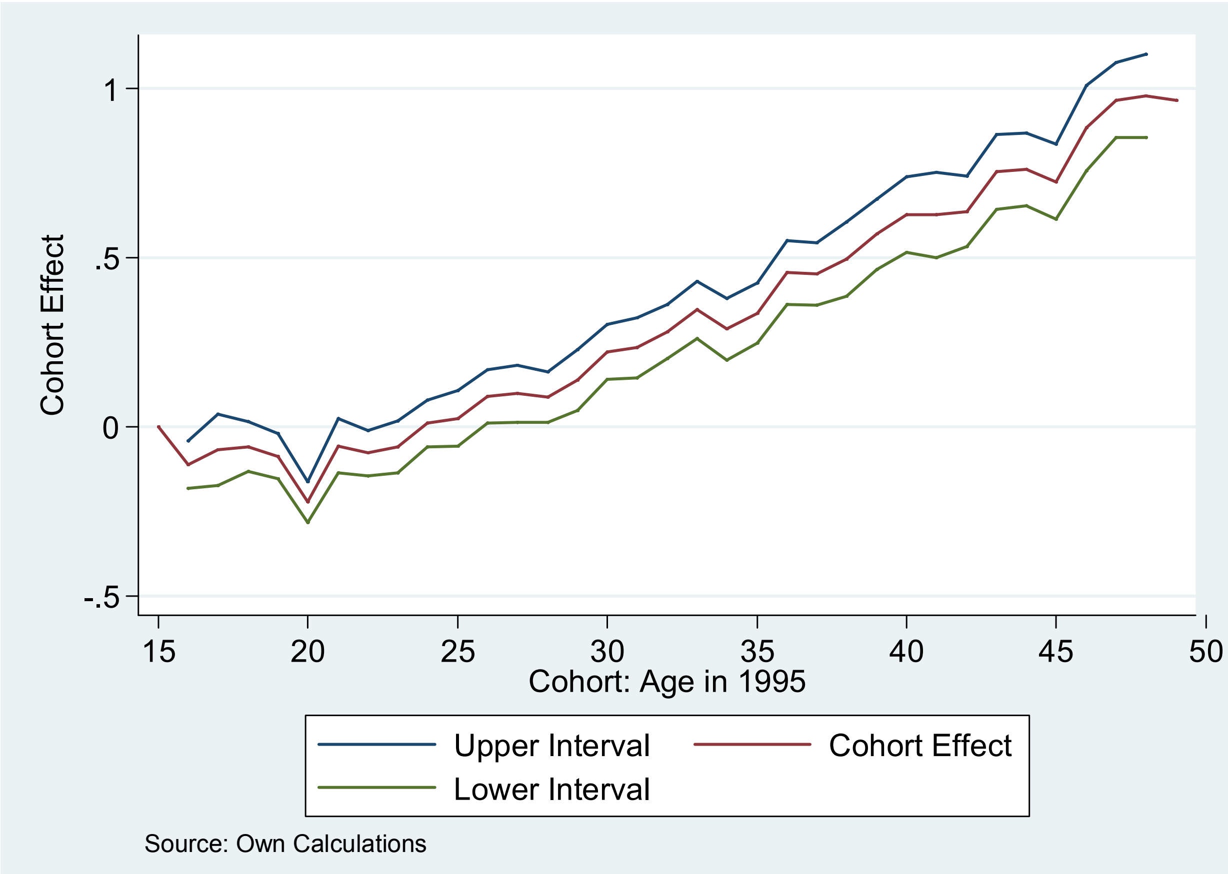


Figure 5: Coefficient Estimates and 95 Percent Confidence Intervals of the Cohort Effects on Marriage Trends



References

Casale, D. (2003). The rise in female labour force participation in south africa: An analysis of household survey data, 1995 – 2001, *School of Economics, University of Natal, Durban* .

Casale, D. and Posel, D. (2002). The continued feminisation of the labour force in south africa: An analysis of recent data and trends, *South African Journal of Economics* **70**(1): 156–184.

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* **74**: 829–836.

Deaton, A. (1997). Analysis of household surveys: A microeconometric approach to development policy, *The Johns Hopkins University Press, Baltimore.* .

Denis, P. and Ntsimane, R. (2004). Absent fathers: Why do men not feature in stories of families affected by hiv/aids in kwazulu-natal, *Paper presented to the Masculinities Symposium, University of Witwatersrand* .

Fienberg, S. and Mason, W. (1978). Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology,* **8**: 1–67.

Firebaugh, G. (1997). Analyzing repeated surveys, *Sage University papers* .

Fu, W. (2000). Ridge estimator in singular design with application to age-period-cohort analysis of disease rates, *Communications in Statistics-Theory and Method* **29**: 263–78.

Fu, W., Hall, P. and Rohan, T. (2004). Age-period-cohort analysis: Structure of estimators, estimability, sensitivity and asymptotics, *Journal of the American Statistical Association, revised and resubmitted* .

Glenn, N. (1976). Cohort analysts futile quest: Statistical attempts to separate age, period, and cohort effects, *American Sociological Review* **41**: 900–904.

Halmiton, L. (2004). Statistics with stata, *Brooks/ Cole-Thomson Learning, Belmont, USA* .

Hobcraft, J., Menken, J. and Preston, S. (1982). Age, period and cohort effects in demography: A review, *Population Index* **48**: 4–43.

Hosegood, V., McGrath, N. and Moultrie, T. (2009). Dispensing with marriage: Marital and partnership trends in rural kwazulu-natal, south africa 2000-2006, *Demographic Research* **20**: 279–312.

Hunter, M. (2005a). Building a home: Unemployment, intimacy, and aids in south africa, *PhD thesis, Department of Geography, University of California at Berkeley* .

Johnston, D. (1981). Measurement of subjective phenomena, *Government Printing Offices, Washington, DC* .

Kalule-Sabiti, I., Palamuleni, M., Makiwane, M. and Amoateng, A. Y. (2007). Family formation and dissolution patterns in families and households in the post-apartheid south africa, *HSRC Press* .

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimations, *Annals of' Statistics* **28**: 1356–78.

Miyazaki, Y. and Raudenbush, S. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design, *Psychological Methods* **5**: 44–63.

O'Brien, R. (2000). Age-period-cohort characteristic models, *Social Science Research* **29**: 123–39.

Papke, L. and Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates, *Journal of Applied Econometrics, Vol.11, No.6, pp619-632.* **11**(6): 619–632.

Rasinski, K. (1988). The effect of question wording on public support for government spending, *NORC, Chicago* .

SADA (2001). Statistics south africa: October household survey, 1999, *South African Data Archive, Pretoria* .

Smith, T. (1987). That which we call welfare by any other name would smell sweeter: an analysis of the impact of question wording on reponse patterns, *Public Opinion Quarterly* **51**: 71–83.

Smith, T. (1988). Timely artifacts: A review of measurement variation in the 1972-1988 gss data (methodological report no. 56), *NORC, Chicago* .

Smith, T. (1992). Some thoughts on the nature of context effects. in n. schwarz and s. sduman (eds.), context effects in social and psychological research, *Springer-Verlag, New York* pp. 163–184.

StatsSA (2001). Marriages and divorces 1999, *Pretoria: Stats SA* .

Tucker, M. and Mitchell-Kernan, C. (1995). Marital behaviour and expectations: Ethnic comparisons of attitudinal and structural correlates. in tucker, m.b. and mitchell-kernan, c. (eds), the decline in marriage among african americans: Causes, consequences and policy implications, *Russel Sage Foundation* .

Wilmoth, J. (1990). Variation in vital rates by age, period, and cohort, *Sociological Methodology* **20**: 295–335.

Wittenberg, M. (2002). Search in south africa: A nonparametric analysis, *The South African Journal if Economics* **70**(8).

Yang, Y. (2006). Bayesian inference for hierarchical age-period-cohort models of repeated cross-section data, *Sociological Methodology* **36**: 39–74.

Yang, Y. (2007). Is old age depressing? growth trajectories and cohort variations in late life depression, *Journal of Health and Social Behavior* **48**: 16–32.

Yang, Y. (2008). Trends in u.s. adult chronic disease mortality, 1960-1999: Age, period, and cohort variations, *Demography* **45**.

Yang, Y., Fu, W. and Land, K. (2004). A methodological comparison of age-period-cohort models: An intrinsic estimator and conventional generalised linear models, *Sociological Methodology* **34**: 75–110.

Yang, Y. and Land, K. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores, *Sociological Methodology* **36**.

Yang, Y., Schulhofer-Wohl, S., Fu, W. and Land, K. (2008). The intrinsic estimator for age-period-cohort analysis: What it is and how to use it, *American Journal of Sociology* **113**.

Appendices

A Nonparametric Approach: LOWESS Technique

LOWESS is a data analysis technique for smoothing scatter-plots, (x_i, y_i), $i = 1, ..., n$ which have been contaminated with noise. Scatter-diagram smoothing involves drawing a smooth curve on a scatter diagram to summarize a relationship, in a fashion that makes very few initial assumptions about the form or strength of the relationship. Cleveland (1979) proposed the LOWESS algorithm as an outlier resistant method based on local polynomial fits.

The basic idea of LOWESS technique is to first fit a polynomial regression in order to find parameter(s) which minimize the influence of outliers on the curve. A fitted value is calculated at each position of the window in x. The technique then moves a window along the x-axis of a scatter-plot. The fitted values at each position of a window, when connected, form the LOWESS smoothed curve.

The fitted value at x_k is the value of a line fit to the data using weighted least squares where the weight for(x_i, y_i) is large if x_i is close to x_k and small if x_i is further away from x_k. Plotting the smoothed points, which form a nonparametric regression of y on x, frequently allows the visual effects on the scatter-plot that are otherwise difficult to detect.

Caution must be taken so that the scatter-plots are not over/under-smoothed. A tradeoff between bias and variance arises with respect to the degree of smoothing such that over-smoothing increases bias but reduces variance while under-smoothing does the reverse. One simple solution towards optimal smoothing is choosing an optimal window within which a point can be smoothed. Within the window defined by the bandwidth, weighted least squares regression of y on x is run using the following "tricube" weights:

$$w_i = \left(1 - |u_i|^3\right)^3$$

where $u_i = \frac{x_k - x_i}{d_i}$ and d_i is the distance between x_k and its furthest neighbour within the band. The weighting allocates a value of 1 for $x_k = x_i$ and declining weights for points further away, reaching zero at the borders of the bands. Thus, points near the middle of the window receive more weight than points towards the boundaries.

For this study, a constant bandwidth of 0.3 is chosen (following Wittenberg (2002)). This indicates that 30 percent of the sample is contained in each band and is used to smooth each point. LOWESS technique thus fits a line to data, portion by portion. The fitted x_i are used as $f(x_i)$ in the model $y = f(x_i) + \varepsilon_i$ where ε_i is assumed to be a white noise process. The result of this process is that the predicted smoothed y values for n observations is obtained by "stitching" together n regressions (Halmiton, 2004). In this case, $f_j(x_i)$ is the proportion in each marital state j; j =married, single, widowed, divorced controlling for age, gender and population group. We will therefore have a nonparametric estimation which will create a scatter of points which follow a distinct trend of plots of proportions on the y-axis of each age category on the x-axis that is in a particular marital status. Such representations will be drawn for the years 1995 to 2006. The slopes of the resulting graphs from the analysis give the estimates of the underlying flow rates, thereby giving information about changing levels in the marital categories over the years.

Specifically interested in levels and flows, this graphical analysis will identify the key turning points and other broad trends in the data. The method is fairly simple and yet yields fruitful results.

B Singulate Mean Age at Marriage

The Singulate mean age at marriage (SMAM) index, originally developed by John Hajnal in 1953, is what is normally estimated in the absence of data on actual age at marriage. Two key assumptions are made in the calculation of SMAM. First, nobody marries before the age of 15, and second, those never married by age 50 will never marry. Proportion never married is then calculated for eight quinquennial age groups in the age range 15-54. These are disaggregated by gender and race. We assume that the proportion never married is P_{gra}, where P is proportion, g stands for gender, r stands for race and a stands for quinquennial age group.

The first stage in the calculation of the SMAM involves calculation of the mean years of singleness lived per hundred persons of the population in question, as follows:

under 15: $15 * 100 = 1500 = Years_{<15}$

15-49: $\sum_{a=15-19}^{45-49} P_{gra} * 5 = Years_{15-49}$

<15 to 50: $Years_{<15} + Years_{15-49} = Years_{<15-50}$

$Years_{<15-50}$ however needs to be adjusted in order to account for the fact that some people will stay unmarried their whole life. This is done by averaging the proportions never married for the two quinquennial age groups, 15-54 as follows:

$\left(\sum_{a=45-49}^{50-54} P_{gra}\right)/2 = A$

It follows, therefore, that the total number of years of singleness experienced by those never married by age 50 is $A * 50$. This implies that for those in the marrying population, the total number of years experienced per hundred person of the whole population is $Years_{<15-50} - (A * 50)$. However, the actual population of people married by age 50, denoted M is $100 - A$. Finally, the SMAM is given by $\frac{Years_{<15-50}-(A*50)}{M}$.

C Other Tables

Table 1: Racial Differentials of Percentage Decline in Marriages, by Gender

	1995	**2006**	**1995**	**2006**	**Male**	**Female**
	Male		**Female**		**% Change**	
All Races	.3741	.3155	.3992	.3596	-15.7	-9.9
	(.4839)	(.4647)	(.4897)	(.4799)		
African	.3231	.2717	.3496	.3149	-15.9	-9.9
	(.4677)	(.4448)	.4768)	(.4645)		
Coloured	.423	.4004	.4233	.4168	-5.3	-1.5
	(.4941)	(.4901)	(.4941)	(.4931)		
Indian	.5349	.5298	.5748	.6158	-0.1	-7.1
	(.499)	(.4996)	(.4946)	(.4868)		
White	.6248	.5444	.6719	.6297	-12.9	-6.5
	(.4842)	(.4982)	(.4696)	(.4831)		

Standard deviations in parentheses

Table 2: Singulate Mean Age At Marriage, by Gender and Race

	Africans		**Whites**	
	Female	**Male**	**Female**	**Male**
1995	29.6	32.3	23.1	25.7
1996	29.3	32.0	24.8	27.7
1997	29.6	32.5	23.8	26.6
1998	29.2	31.9	24.3	27.7
1999	29.4	31.2	25.0	27.7
2000	29.8	32.8	24.8	27.4
2001	30.3	33.0	25.0	27.7
2002	29.9	32.8	25.0	27.5
2003	30.5	33.3	25.5	28.0
2004	30.4	33.6	25.4	28.2
2005	30.8	33.7	26.3	27.9
2006	31.0	34.0	25.8	28.9

Table 3: Cohabitation Factor among Africans

	1995	**1996**	**1997**	**1998**	**1999**	**2004**	**2005**	**2006**
Married	.3057	.3167	.2886	.2746	.2746	.2318	.2105	.2036
	(.4607)	(.4652)	(.4531)	(.4463)	(.4427)	(.422)	(.4077)	(.4027)
Cohabiting	0439	.0477	.0539	.0726	.0649	.106	.1095	.1113
	(.2049)	(.2131)	(.2258)	(.2595)	(.2464)	(.3079)	(.3122)	(.3145)

Standard deviations in parentheses