Social Interactions, Ethnicity and Fertility in Kenya*

Sriya Iyer
Faculty of Economics
University of Cambridge

Melvyn Weeks
Faculty of Economics
University of Cambridge

January 2009

*Acknowledgements: For helpful comments and discussions we are grateful to Wiji Arulampalam, Ethan Cohen-Cole, Partha Dasgupta, Gernot Doppelhofer, Steven Durlauf, Timothy Guinnane, Andrew Harvey, Larry Iannaccone, Hashem Pesaran, Richard Smith, and Chander Velu. We acknowledge funding from the Centre for Research in Microeconomics, and St. Catharine's College, Cambridge. We are especially grateful to Mrs. Zipporah Onchari and the Director of the Kenya Meteorological Services for giving us access to their rainfall data and for patiently answering our many questions. We are indebted to Anne Mason, Chris Brown and Tirthankar Chakravarty for excellent research assistance. We are also grateful for many helpful suggestions provided by seminar participants at George Mason University, University of Wisconsin-Madison, University of Warwick, the School of Oriental and African Studies, University of London, and the Brookings Institution.

Abstract

This paper examines, theoretically and empirically, the impact of reproductive externalities on fertility behaviour in Kenya. We examine this issue by identifying structural forms of social interaction operating across individuals belonging to different ethnic groups on the number of children ever born. We use the 1998 Demographic and Health Survey, and meteorological data on Kenya, to examine whether social interactions effects are important over and above an individual's characteristics in order to explain variations in fertility. In so doing, we conclude that social interactions are very important for the fertility behaviour of different ethnic groups in Kenya.

1 Introduction

Both economists and demographers have examined the balance between economic and non-economic factors in orchestrating a fertility transition, as witnessed in historical European populations and in some East Asian economies over relatively short periods of time. Classic studies of fertility such as the Princeton Fertility Project highlighted that the transition to low fertility in historical European populations occurred in a variety of socio-economic and institutional contexts, with a significant role being played by the local social environment (Coale and Watkins (1986), Federici, Mason, and Sogner (1993)). More recent studies have emphasised the role of the social climate and the influence of social interactions on demographic behaviour (Hank (2001), Manski and Mayshar (2003), Munshi and Myaux (2006). These empirical findings have led many economists naturally to focus on modelling the influence of social interactions on contemporary fertility transitions (Durlauf and Walker (1999); Manski (1993), Manski and Mayshar (2002)) and on the reproductive externalities and coordination failures associated with fertility behaviour per se ((Dasgupta (2000); Kohler (2001); Iyer and Velu (2006)).

Empirical and theoretical analyses of social interactions can be divided into two broad approaches. One approach considers the impact of social interactions within predetermined groups (Akerlof (1997); Brock and Durlauf (2000), Brock and Durlauf (2001)), and emphasises how interactions affect individual and group-level outcomes in a cross-section. A second approach analyses how group formation results from social interactions with particular emphasis on the growth of residential neighbourhoods. This literature has focused on questions of geographic proximity and dynamic group formation (Borjas (1995); Ginther, Haverman, and Wolfe (2000); Conley and Topa (2002)).

Our study is located within the first strand of the literature and examines social interactions in the context of fertility behaviour in Kenya where groups are considered predetermined. We conceive of the fertility of a Kenyan woman to be influenced by a range of factors such as her individual characteristics and the characteristics of the household to which she belongs. We postulate that the social environment impinges on a woman's behaviour through the multiple interactions that she undertakes both at the household level and at the level of other reference groups. In doing so we highlight a crucial point often overlooked in both the theoretical and empirical literature on this issue: namely that there are different levels at which social interactions occur. In doing so we consider the possibility that both *local and global forms of interaction* may coexist and be important, characterised by different mechanisms of social interaction (Horst and Scheinkman (2003a)). We argue that this has profound implications, both theoretically and empirically, for studies of fertility behaviour and social interactions, and for the study of social interactions and economic phenomenon, more widely. The research presented in this paper quantifies the effects of social interactions and group membership on the number of children ever born. In this respect our work highlights the existence of uncertainty as to whether these effects are mediated through household composition, ethnic affiliation, a neighbourhood cluster effect, or some combination thereof. There is thus uncertainty both with respect to the levels at which interactions occur, and the expressions of structural dependence at each level.

Section 2 summarizes the literature on social interactions and fertility behaviour. In section 3 we provide an overview of ethnic groups in Kenya, highlighting the characteristics which are relevant for fertility behaviour. In section 4 we present a model of reproductive externalities with social utility effects. In sections 5 and 6 we consider, respectively, econometric models of fertility bahaviour and the specification of the conditional mean. Sections 7 examines the question of identification and discusses how we use unique district-level rainfall data from Kenya's meteorological rainfall stations over time to identify variations in fertility by ethnic group. In Section 8 we present the demographic and meteorological rainfall data. We discuss our results in section 9, and section 10 concludes.

2 Social Interactions and Fertility

The conventional economic household demand model of fertility behaviour posits that a couple's fertility is a function of the money costs of children and the opportunity costs of the value of parental time (Becker (1981)). Conditional upon identifying a fixed set of fundamentals as comprised of a vector of characteristics of both the individual and the household, an atomistic model of fertility behaviour simply focusses on the direct affects attributable to a change in these fundamentals. Subsequently, if we then observe that the variance in fertility outcomes is in excess of that which might be accounted for by differences in fundamentals,[1] an extension of the standard fertility model to allow for the existence of social multiplier effects would seem reasonable (see, for example, Becker and Murphy (2000)).

Theoretical work of Dasgupta (2000) and Kohler (2001) has argued that a couple's fertility may be influenced by the level of fertility of all other couples within a society. In the most general sense, this is what is meant by social interaction in a fertility context: it is the public interaction between individuals in a society as they perceive each other and observe each others' fertility behaviour. This consequently alters their social environment, which in turn ultimately influences their private decision-making about fertility.

[1]Horst and Scheinkman (2003b) note that any change in the fundamentals exerts direct effects and indirect effects that have the same sign. They comment that when the indirect effects are very significant, the multiplier is very large.

As a consequence, the existence of social interactions may lead to multiple equilibria and coordination failures in demographic decision-making, shown by a high level of variability in outcomes for a given set of fundamentals. Recent economic analyses of fertility behaviour have been much concerned with social interactions (see Montgomery, Kiros, Agyeman, Casterline, Aglobitse, and Hewett (2001), Hank (2001), Nauck (1995), Becker and Murphy (2000), Kohler (2001), and Manski (2000), Munshi and Myaux (2006)).

In the social interaction literature[2] and following Manski (1993), three hypotheses are often advanced to explain the observation that individuals belonging to a common reference group tend to behave similarly, even after controlling for a set of observed individual characteristics. In a fertility context, there is an endogenous effect if, ceteris paribus, a woman's children ever born (CEB) tends to vary with the average CEB of members of her ethnic group or locality, perhaps because she suffers a utility loss from deviations from group behaviour. There is an exogenous or contextual effect if, ceteris paribus, a woman's CEB tends to vary with (or more plausibly inversely with) the average educational attainment of her ethnic group or locality. For example, the existence of educated neighbours may foster positive attitudes towards smaller family size. Finally, there is a correlated effect if, ceteris paribus, women in the same ethnic group or locality tend to have similar CEB because they are, for example, similarly wealthy. The importance of differentiating between these three hypotheses can be seen by examining the respective policy implications. Consider, for example, an intervention to provide free contraception to some members of an ethnic group or a number of neighbourhoods. If there are endogenous effects, then an effective policy of free contraception may both directly reduce the fertility of the recipients but, as their fertility decreases, indirectly reduces the fertility of all other members of that ethnic group or neighbourhood, with a feedback to further fertility reductions by the recipients of the free contraception. Exogenous effects and correlated effects will not, in general, generate this kind of social multiplier effect.

It is possible to conceive of a large number of exogenous effects; in the specific context of fertility behaviour, education plays a particularly prominent role. The effect of social interactions on fertility behaviour operates through education at multiple levels. For example, the potential effects of education on fertility can be decomposed into 2 components. The first is the effect of individual educational attainment on individual-level fertility. In a developing society, this would be dependent on the money-costs of acquiring an education and the opportunity costs of wages foregone. Second, there is also an iconic value of education in the sense that some individuals may aspire to the attributes of other higher educated groups in a population. In some developing societies, the iconic value of an education is very high. For example, sociologists of India often comment on the phenomenon of 'Sanskritization' in which lower-caste groups take on the

[2] For an excellent discussion of the theoretical literature on economic models of social interactions and the influence of non-market institutions on market institutions see Glaeser and Scheinkman (1999)).

characteristics and customs of the upper castes in order to gain greater legitimacy and status in the Indian social system. Frequently this manifests itself in the desire to acquire an education, or to continue one (Srinivas (1994)).

It is also important to make the distinction between expected versus observed behaviour in locating the mechanism by which a group impacts upon individual decision making. One of the central issues here is the notion that in large groups, individuals do not observe the behaviour of group members, and therefore the appropriate object is to consider interaction effects as mediated by subjective expectations. However, such an argument does not hold in cases where the size of the group is small, such as interaction effects within extended families of small communities.

3 Overview of ethnic groups in Kenya

Social interactions and channels of message transmission about fertility behaviour are important at the level of ethnicity. Members of different ethnic groups speak the same language, usually adopt similar cultural practices, and with some exceptions in border areas, reside closely in the same, or in contiguous districts (Fapohunda and Poukouta (1997)). In the specific case of Kenya it has been argued that cultural norms may encourage high fertility but also contribute to fertility decline over time (Caldwell and Caldwell (1987); Ascadi, Ascadi, and Bulatao (1990)). We argue that individuals' multiple levels of social interactions reflect their social identities - regional identities, ethnic identities, religious identities, linguistic identities. While we do not explore the question of identity in this paper, we do focus on how local and global interactions that arise from these identities, significantly affect individuals' behaviour. Our analysis is based upon historical and anthropological studies of Kenya with particular focus on the features of Kenya's major ethnic groups that have relevance for fertility behaviour. In this section we discuss the ethnic groups with reference to their population composition, the region of their residence, and three characteristics of these groups which are noteworthy - their residential pattern of settlement, clan organisation and education. These characteristics have implications for the politics of ethnicity and fertility in Kenya.

Population The population of Kenya consists of three main groups - the Africans, Asians and Europeans. Although the last two groups mainly reside in towns, 90% of the African population continues to live in rural areas (see Meck (1971)). The tribal groups of rural Kenya live in clearly defined settlements in the more remote areas. In the Lake Victoria basin, the highlands and the coast, there is a more heterogeneous population structure (Morgan and Shaffer (1966):2; Meck (1971): 24). The tribes can be defined broadly into four groups as classified by their language: the Bantu, the Nilotic, the Nilo-Hamitic, and the Hamitic. Within these four broad language groups, ethnic groups represent an additional sub-division. According to the 1969 *Population Census* there are 42 different ethnic African groups in Kenya (Rep (1970)). The five largest groups, accounting for over 75% of the population, are the Kikuyu (22%), Kamba (11%), Kalenjin (12%), Luhya (14%) and Luo (18%).[3] The Kisii and the Meru both comprise 6% of the total population. Other groups, such as the Maasai and the Somali constitute 15% of the population.[4] A religious breakdown of the ethnic groups show that most Kikuyu are Christians, probably because they were the Kenyan tribe with the closest links with Christian missions historically (Meck (1971): 27). Over 90% of Luhya and Luo are Christians, compared to over 60% of Kamba. Some of the Mijikenda sub-groups, such as the Digo, are Muslim.

Census data from Kenya depict a steady increase in the total population of Kenya since 1948 to the present. The total fertility rate in Kenya has fallen from an average of 6 births per woman in 1948 to approximately 4.5 births today (Ajayi and Kekovole (1998)). However, the most significant declines in fertility have occurred in the last twenty years: the DHS studies conducted in 1989, 1993 and 1998 showed that the TFR dropped from 6.7 in 1989 and 5.4 in 1993 to 4.7 in 1998. This drop in fertility is considered one of the most dramatic recorded anywhere in the developed and developing world (Ajayi and Kekovole (1998): 116). The fertility differences by ethnic group are also very large, as shown in Table 1. For women aged 15-49, the mean CEB varies from 2.5 for the Kikuyu to 3.5 for the Luo. Previous studies of fertility and ethnicity in Kenya have revealed that desired family size is smaller among the Kamba, Kikuyu, Luhya compared to the Luo (Fapohunda and Poukouta (1997)). The Luo exhibit a strong preference for large families and it is reported use contraception much less than other groups (Watkins, Rutenberg, and Green (1995)).

Region Ethnic boundaries in Kenya, to a very large extent, are coterminous with political and administrative boundaries (Fapohunda and Poukouta (1997)). Today the regional breakdown of the different ethnic groups is as follows: the Kalenjin reside in the Rift Valley; the Kikuyu live in the Central region, but have also migrated to Nairobi and the Rift Valley. The Meru/Embu reside in the North and East. The Luhya live in the Western province, but have also migrated to Nairobi and Mombasa. The Luo live in Nyanza, with Kisumu as their capital, but have also migrated to the Rift Valley. The Kamba live in close proximity to Nairobi and exhibit ethnic affiliation to the Kikuyu. The Mijikenda/Swahili live in the Coast province. The Meru and Embu groups neighbour the Kikuyu to the north and east. The Luhya, who live in Kenya's Western province, are a less homogenous group compared to the Kikuyu and Kamba (Were (1967)). Among the

[3] According to the most recent Census.

[4] The Asians, Europeans and Arabs make up about 1% of the population.

Luhya, there is a distinction that needs to be made between those in the north and those in areas such as Kakamega, with high population density. Although the Luo who live in Nyanza depict a rate of population growth almost as high as the Luhya, they are a more urbanised people (Ominde (1968)). For the Kamba, proximity to Nairobi and their close ethnic affiliation to the Kikuyu are particularly significant (Berg-Schlosser (1984). The Mijikenda and Swahili groups who reside in Kenya's Coast province are the most agricultural of all of Kenya ethnic groups.[5]

Residential pattern of settlement One of the most notable features of these ethnic groups is their pattern of residence.[6] It is important to emphasise that the notion of the 'village' in this society is rather more diffuse than in the compact settlements of Europe and elsewhere. Although there are differences between ethnic groups, families are grouped mainly in homesteads that are located in clusters, and which form the basis of community interaction. In some of these clusters, for example, among the Kikuyu there exist groups of households who are not merely coresident, but also related by blood and marriage. In other clusters, there are households that are located in close proximity, but where individuals in these households are not related to other individuals in neighbouring households.

For example, the Kalenjin live mainly in homesteads that are individual family based. Property is usually inherited along paternal lines or male agnates. The smallest unit of territorial composition in Kalenjin society is the '*temenik'* or hamlet, a cluster of homesteads; these are usually grouped into 'villages' of 15 to 60 *temenik*. The Kikuyu live in homesteads on land owned by the family, and surrounded by fields (or '*shamba*'). Traditionally land was owned by many households which constitute the extended family located on the same 'ridge' which is the traditional geographical division[7] (Berg-Schlosser (1984): 50). Supplementing the organisation of Kikuyu society based on kinship, there is also a geographical demarcation. The '*itura*' or village consists of groups of families. This is further subdivided into those living on the same ridge. The pattern of Luo settlement is similar to other communities, although their homesteads usually consist of families who own land, dispersed as a safeguard against climatic conditions. The Kamba are also agriculturists, and their pattern of settlement is similar to the Kikuyu (Middleton and Kershaw (1965)). The traditional land-owning unit is the extended family. In terms of territorial residence, the most basic unit is the *ukambani,* which is a homestead that comprises several extended families. Several of these are grouped to form a *kivalo* within

[5] The word 'Mijikenda' means 'nine towns' or 'tribes'. The Mijikenda consist of nine distinct subgroups - Giriama, Duruma, Digo, Rabai, Chonyi, Kambe, Kauma, Ribe, and Jibana.

[6] A key point to note here is that villages are not defined units of settlement, although their boundaries are well-known locally. These boundaries are usually marked by trees, stones, and so forth (Berg-Schlosser (1984): 139)

[7] This was altered with the land reforms of 1950 when ownership of land was transferred to individuals.

which social interactions, especially marriage, take place. For the Mijikenda, the main form of residence is the *mudzi* or village which consists of groups of agriculturists and fisherman.

Clan organisation These tribal groups are also broadly characterised by three main features that have relevance for social interactions: the importance of the family group, the clan, and the system of age-grading[8] (Meck (1971)). For example, anthropological studies of the agriculturalist Kikuyu argue that among them, fathers exert a less important role for economic decision-making, that relations between the extended family are strong, and that they exhibit a highly evolved sense of ethnic identity which can often override more national concerns (Berg-Schlosser (1984), Ferguson and SrungBoonmee (2003)). Anthropologists comment that among the Kikuyu, the most economically and socially effective unit is the *mbari*: 'a group of families who trace their descent from a common ancestor following the paternal line, often for up to seven or eight generations. In addition to its functions as the most important traditional land-holding unit in Kikuyu society, the *mbari* is an important social reference-group for many Kikuyu and still plays an effective role in many economic and social relationships including the more modern ones.' (Berg-Schlosser (1984): 53). We would therefore expect that the endogenous social interaction effects among this group would emerge as being particularly strong. The Luhya do not have as powerful a clan organisation as some of the other groups (Meck (1971): 28). Anthropologists comment that a feature of the Luhya that makes them quite distinct from other groups is the relatively strict accepted norms of behaviour, particularly with respect to marriage and interactions between the sexes (Berg-Schlosser (1984): 114).

Education Initiated first by Jomo Kenyatta, education policy in Kenya has been pursued actively by the government, with remarkable success, especially with increases in primary and secondary school enrolment for girls. However, despite the general increases in the uptake of education and policies such as school fee remission and the development of *harambee* or self-help community schools which foster these, the mean number of years of schooling differs considerably by ethnicity.

The mean number of years of education varies considerably by ethnic group in Kenya, as shown in Table 2. This factor has relevance for social interactions as we might expect the importance of social interactions on fertility to be mediated by the effect of education. Previous studies of Kenyan fertility argue that the Kikuyu and the Kamba show the least preference for large families because they had early access to colonial education

[8]The system of age-grading is a form of vertical stratification which every member of a tribe goes through during the course of their lives, with each 'age-grade' made distinct from the other with certain rites of passage (Kenyatta (1961):2; Berg-Schlosser (1984): 55).

(Fapohunda and Poukouta (1997)). The Kikuyu are the best educated, compared to other ethnic groups in Kenya (Berg-Schlosser (1984): 60; De Wilde (1967): 39). We can explain this development historically - many Kikuyu worked as wage labourers in European-owned plantations and attended schools. As a community, they viewed education as the means to progress and this led them to acquire positions of responsibility in the colonial administration. In the post-Independence period, the Kikuyu continued to dominate and consolidate their position, politically and economically, relative to the other groups (Ferguson and SrungBoonmee (2003)). The Luhya populations, are also highly literate groups. Most Kamba are enrolled in formal schooling. Among the Kalenjin, levels of education are, in general, less than other groups. The Mijikenda and Swahili groups also depict low levels of education compared to other groups in this population.

Politics of ethnicity Kenya has been much in the news recently because of recent events which have highlighted the importance of the politics of ethnicity in this country. Recent evidence has highlighted that differences in the economic performance of ethnic groups can be traced to their development historically and to economic decisions made by these groups (Ferguson and SrungBoonmee (2003)). Ethnic identity in Kenya has always been very strong historically, and it continues to be a very potent force in Kenya's politics even today. For example, the Kalenjin, who are a heterogenous ethnic group and who live primarily in Kenya's Rift Valley province (Were (1967)) are closest to Nairobi geographically, and they are politically and economically Kenya's dominant ethnic group (Kenyatta (1966)). Their location in the eastern Rift Valley allows them to enjoy a level of political proximity that, in part, determines their (relatively) superior economic status. Agricultural innovation first arose among the Kikuyu who also absorbed land reform more readily than other groups (Meck (1971): 27). Historically, the Kikuyu were one of the first ethnic groups in Kenya to absorb European-style capitalism in the form of wage labour and participation in the monetary economy, so in contrast for example to the Maasai whom they neighbour, the Kikuyu's subsequent economic performance has been much better (Ferguson and SrungBoonmee (2003)). [9] In the past the Luo were less responsive to social and economic change compared to other tribes (Meck (1971)). For example, land reform met with a great deal of resistance among this group as it was believed to conflict with religious beliefs. But since 2001, the Luo have increasingly been incorporated into the government. In more recent times, relations between the ethnic groups has been closely tied with the development of Kenyan politics, particularly since the introduction of multiparty politics in Kenya since 1991 after over twenty years of rule by one party (Throup (2001)). The increased strength and importance of the ethnic

[9] For example, following the Land Transfer Programme, Kikuyu land owners in the Central province specialised in the production of coffee which is widely exported by Kenya, earning these farmers resources that ensured their success as a community (Ferguson and SrungBoonmee (2003)).

identity has also led, more worryingly, to ethnic clashes in 1992-93 and in 1997.

So ethnicity - or ethnic group identity - is important for this society. Ethnicity has displayed an immutability which may be accounted for by the historical evolution of these population groups, manifest in the recent political history of this nation.

4 A Model of Fertility with Social Interactions

We consider a population in which there is a representative household h[10] characterised by a location j, a type s - here ethnicity - and a vector of characteristics $\mathbf{x}_{i.}$. We assume that both j and s are fixed for all individuals, with characteristics potentially varying across individuals, type and location. Our point of departure is the standard Beckerian utility function (Becker (1981), Willis (1973))

$$U_i = U(n_i, \mathbf{z}_i, \mathbf{x}_i, \boldsymbol{\varepsilon}_i), \tag{1}$$

where U_i denotes the utility function of household i, n_i represents the number of children, $\mathbf{z}_i$ is all sources of satisfaction to the husband and wife other than those arising from children, and $\mathbf{x}_i$ denotes the vector of socio-demographic characteristics which affect preferences. By including the stochastic term ε_i we allow for imperfect information on the part of the analyst. This is subject to the usual budget constraint:

$$I = n_i p_n + \mathbf{z}_i \mathbf{p}_z, \tag{2}$$

where we assume that total income I is expended on children n_i and on all sources of satisfaction other than those arising from children $\mathbf{z}_i$; p_n and $\mathbf{p}_z$ are the shadow prices of children and other sources of satisfaction respectively.

In extending the above atomistic utility model to incorporate a theory of social interactions we first write

$$V_i = (U(n_i, \mathbf{z}_i, \mathbf{x}_i); S^j(n_i, \mathbf{n}_{-i}^j), S^s(n_i, \mathbf{n}_{-i}^s), \boldsymbol{\varepsilon}_i), \tag{3}$$

where V_i, total individual utility, is comprised of a private utility term $U()$, and two social utility terms: $S^j(n_i, \mathbf{n}_{-i}^j)$ represents a social utility term for location j, given by some function of the number of children choices made by all other residents of j; $\mathbf{n}_{-i}^j = (n_{1,}^j n_2^j, ..., n_{i-1}^j, n_{i+1}^j, ..., n_{M_j}^j)$, where M_j denotes the number of individuals in location j. $S^s(n_i, \mathbf{n}_{-i}^s)$ represents a social utility term for type s, namely the impact of some

[10] Note that within the household we consider husband's and wife's preferences to be coincident. In this respect we do not examine intra-household bargaining between the couple in the determination of fertility outcomes. For theoretical simplicity, we assume that the 'household' represents the couple. Empirically this can be translated into a measure of the woman's fertility, measured by the total number of children borne by her.

function of the number of children choices made by all other members of type s, where $\mathbf{n}^s_{-i} = (n^s_{1,} n^s_2, ..., n^s_{i-1}, n^s_{i+1}, ..., n^s_{M_s})$. In both cases the extension allows us to capture that component of individual utility attributable to a fertility choice n_i, that is dependent upon the fertility choices of others. The 'others' in question may be those, for example, resident in the individual's household, or others resident in the local community for example, in the individual's street or neighbourhood. Rewriting (3) as additive in individual and social utility terms, we have

$$V_i = U(n_i, \mathbf{z}_i, \mathbf{x}_i) + S^j(n_i, \mathbf{n}^j_{-i}) + S^s(n_i, \mathbf{n}^s_{-i}) + \varepsilon_i. \tag{4}$$

In extending (1) by adding social utility functions $S^j(.)$ and $S^s()$, our intention is to develop a model which facilitates a distinction between *local* and *global* interactions in the sense that individual behaviour may also be determined by the decisions of a larger ethnic group. In referring to a local interaction, individuals have incentives to conform to the behaviour of a small number of appropriately defined neighbours, whose actions they can 'observe'. In contrast a global interaction refers to a situation where individuals face incentives to conform to the 'expected' behaviour of a common reference group (Horst and Scheinkman (2003b)), whose behaviour we cannot observe. What is significant here is that both a local and a global interaction can occur simultaneously. This specification in (3), including more than a single social utility term, allows for *multiple social interactions* - namely the possibility that social interaction is both localised within a specific location, and more diffuse across a larger, geographically diffuse sub-population classified by type. Such a formulation allows us to examine, for example, whether after controlling for social interaction effects accruing through geographical proximity, there is a residual effect due to the desire to conform to a set of behaviours ascribed by ethnicity.

In focussing on that part of total utility attributable to social utility functions $S^j(.)$ and $S^s(.)$, it is obvious from (4) that cross-partial effects will be key objects. Differentiating the Beckerian first-order conditions with respect to social interactions yields the second-order conditions

$$\kappa^j_{il} = \partial^2 S^j(n_i, n_l)/\partial n_i \partial n_l > 0, \tag{5}$$
$$\kappa^s_{il} = \partial^2 S^s(n_i, n_l)/\partial n_i \partial n_l > 0, \tag{6}$$

where κ^j_{il} represents a measure of the disutility accruing to i from deviating from the behaviour of l for $i, l \in j$. In examining (5) and (6) we define strategic complementarity as representing the increasing marginal utility of woman i as a direct result of the fertility choices of other women in the group. The precise form of interaction will obviously depend on the functions S^j and S^s.

If we are willing to assume that social utility exhibits strategic complementarities that are totalistic and constant (see Cooper (1988)), then we may impose a number of restrictions. For example, the restriction $\kappa_{il}^{j} = \kappa^{j}/2(M_j - 1)$ is consistent with a model of uniform *local* interaction, with equal weights assigned to all members of location j. We note that although such a restriction may be theoretically difficult to justify, it is not possible to identify separate measures κ_{il}^{j} if there is no information as to the relative location of all i, l within each location. Such a restriction implies that, in the case of type, all individuals are equally influential with respect to fertility decisions; and, in the case of spatially defined clusters, all individuals are located at the centroid.

We consider two forms of social utility which are consistent with totalistic and constant strategic complementarity. At the level of location j these may be written as

$$S_1^j(n_i, \overline{n}_{-i}^j) = \theta_1 n_i \overline{n}_{-i}^j \tag{7}$$

$$S_2^j(n_i, \overline{n}_{-i}^j) = \frac{-\theta_2}{2}(n_i - \overline{n}_{-i}^j)^2. \tag{8}$$

(7) represents a proportional spillover form of dependence in the sense that for $\partial S_1^j(.)/\partial n_i = \theta_1 \overline{n}_{-i}^j$, strategic complementarities are solely dependent upon the mean level of children ever born within location j. (8) represents social utility as a measure of conformism with $\partial S_2(.)/\partial n_i = \theta_2(\overline{n}_{-i}^j - n_i)$. Under (8) there is an incentive to conform to mean fertility behaviour within the group with deviations from the mean penalised more severely, relative to the proportional spillovers case. If behaving like others confers additional status on a woman, she may desire to conform.[11] Note also that although

$$\frac{\partial^2 S_l^j(.)}{\partial n_i \partial \overline{n}_{-i}^j} = \theta_l,\ l = 1, 2,$$

such that strategic complementarities are captured by a single parameter in both cases.

We note that in anticipating subsequent data constraints we have represented the social utility function in a way which assumes that all interactions within a group have equal weight. However, the additional restriction on the social utility terms (as in 7 and 8) that any estimated endogenous interactions are constant over a population is unlikely to hold, and is not required for identification. For example, in the case of effects which stem from a desire to conform to a group fertility norm, such a norm may vary *across* different ethnic groups. In the case of the Kikuyu, which reside in clusters comprised of an extended family, then we have prior information suggesting the (ceteris paribus) likelihood of greater interaction than other groups.

[11] There may also be a compelling reason to conform if such status in the community bears economic benefits, e.g. in the allocation of local resources.

4.1 Measuring Social Interactions: The Choice of Reference Groups

Given that our data contain no information which points to the existence of specific groups of individuals which define a network within which social interaction takes place,[12] we acknowledge uncertainty as to what constitutes an appropriate reference group with respect to fertility decisions. However, using a number of anthropological sources we have identified that within rural populations in Kenya, fertility decisions are conducted within families which are members of ethnic groups and which reside mainly in homesteads that are located in relatively small clusters. These various allegiances, we postulate, then form the basis of social interaction. Subsequently we emphasise the role of two reference groups. First, *ethnicity* postulates that the status of women and attitudes towards children may differ substantially across ethnic groups, with differential evaluation of the psychological costs and benefits of bearing children.[13]

Second, we have noted that in the case of Kenya, situated between the individual and the ethnic group, are relatively small *clusters of households* within which physical proximity dictates that individuals directly observe and bear the costs of the decisions of others. In addition we have information which allows us to distinguish between clusters which, for certain ethnic groups, are comprised of households which are blood related, with obvious ramifications for interaction.[14] For example, in the case of the Kikuyu we might expect the extent of such interaction to be on average higher given that clusters of households are generally comprised of groups of individuals which are related by blood and marriage (the *mbari*). In such a situation one might expect to find a stronger normative influence relative to clusters comprised of households, as in the case of the Kalenjin, who live mainly in homesteads that are individual family based.

5 Econometric Models of Fertility Behaviour

In motivating a modelling strategy we first consider the observed data. For each i^{th} woman we observe $\{C_i, \mathbf{x}_i, \mathfrak{e}_i, L_{ij}\}$, where $C_i \in \mathfrak{c} \subseteq \{0, 1, 2, ...\}$ represents a count of the total number of children ever born, $\mathbf{x}_i$ is a vector of characteristics which includes individual characteristics, together with ethnic, cluster, and household attributes, $\mathfrak{e}$ is a categorical variable indicating group membership, and $L_{ij}, j = 1, ..., J$ denotes the j^{th} location in which i resides. At the outset we anticipate a number of econometric issues

[12] For an example where such data exists, see Montgomery, Kiros, Agyeman, Casterline, Aglobitse, and Hewett (2001)

[13] There has been a great deal of research on the effects on fertility of religion and ethnic group membership (see Iyer (2002) for a detailed discussion of this literature).

[14] In other contexts the locus of interaction might consist of a well-defined set of households who live within a given geographical space such as a street (see, for example, Guinnane, Moehling, and O'Grada (2001)).

given that (i) count data is both integer valued and heteroscedastic by construction,[15] (ii) the nature of the interactions across individuals will generate, by construction, a problem of endogeneity, and (iii) dependent upon the information contained in $\mathbf{x}_i$ interaction across $i \in L_j$ is likely to generate non-independence across the error terms.

Below we consider these issues in the context of a baseline Poisson model.

5.1 Departures from Poisson

The Poisson distribution for the number of children ever born $C = 0, 1, 2, ...$, is given by

$$f(C_i|\mathbf{x}_i) = \frac{e^{-\kappa_i(\boldsymbol{\beta})}\kappa_i(\boldsymbol{\beta})^{C_i}}{C_i!}, \quad C_i = 0, 1, 2, ... \tag{9}$$

Given the nonnegativity constraint, the Poisson density in (9) is often accompanied by an exponential conditional mean function $E(C_i|\mathbf{x}_i) = \kappa_i(\boldsymbol{\beta}) = \exp(\mathbf{x}_i'\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ denotes a vector of parameters. Although in modelling the determinants of the number of children ever born to a woman, the Poisson density represents a natural benchmark, the equidispersion property, namely that $E(C_i|\mathbf{x}_i) = Var(C_i|\mathbf{x}_i) = \kappa_i(\boldsymbol{\beta})$ is generally restrictive.

In the baseline Poisson regression model the maintained assumption of the equality of the mean and variance has particular implications for our study. Specifically this implies that the randomness of observations on fertility counts with the same value on the explanatory variable(s) can be described by the same Poisson distribution (i.e., the same mean). This condition may be violated when there exists a random effect for each individual and/or there exists a tendency for observations to cluster. At this juncture it is worth noting that across our sample of married women in Kenya, the mean number of children is 3.19 with the variance approximately three times the mean (9.38).

In considering a number of more flexible specifications we first write the conditional mean function as

$$\kappa_i(\boldsymbol{\beta}, \omega_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}+\omega_i) \tag{10}$$

$$= e^{\mathbf{x}_i'\boldsymbol{\beta}}e^{\omega_i}. \tag{11}$$

(11) accommodates unobserved heterogeneity with the conditional mean represented as a random variable determined by observed characteristics $\mathbf{x}_i$, and variation in a multiplicative i.i.d unobserved heterogeneity component, ω_i. We note that the multiplicative model handles regressors $\mathbf{x}_i$ and error terms ω_i in a symmetric fashion and, in addition, is congruent with both the Poisson and Generalised Linear Model (GLM) formulation given that from (10) and for C not equal to zero, $\ln C_i = \mathbf{x}_i'\boldsymbol{\beta} + \omega_i$, providing a link

[15] This follows given that for any process bounded at zero, the variance will be an increasing function of the mean.

to ordinary least squares. In addition, and as noted by Windmeijer and Santos-Silva (1996), although multiplicative and additive models are observationally equivalent when only the first order conditional mean is specified, there are differences in the presence of endogeneity and the subsequent choice of instruments in the two specifications.

Since ω_i is unobserved, (10) suggests a random-effects interpretation of this particular extension. Conditional on ω_i, the CEB count C_i is distributed Poisson

$$f(C_i|\mathbf{x}_i,\omega_i) = \frac{e^{-\kappa_i(\boldsymbol{\beta},\omega_i)}\kappa_i(\boldsymbol{\beta},\omega_i)^{C_i}}{C_i!}, \quad C_i = 0,1,2,..., \tag{12}$$

with the unconditional probability given by

$$f(C_i|\mathbf{x}_i) = \int_0^\infty [e^{-\kappa_i(\boldsymbol{\beta},\omega_i)}\kappa_i(\boldsymbol{\beta},\omega_i)^{C_i}/C_i!]g(\omega_i)d\omega_i. \tag{13}$$

(10) and (12) can then be used to motivate a class of mean-variance models with the conditional mean nonnegative and the distribution reflecting a mean-variance relationship of the form $Var(C|\mathbf{x}) = \alpha \cdot E(C|\mathbf{x})^\gamma$, where α and γ are unknown parameters. The model is operationalised by either choosing a functional form for ω_i, or adopting a semiparametric estimator. Based on conjugacy, a common *fully* parametric specification utilises a gamma distribution for the mixing distribution,[16] $g(\omega_i)$, with mean 1 and variance α. The resulting mixture of Poisson and gamma components can be interpreted as the Negative Binomial model: the first two moments are $E(C_i|\mathbf{x}_i) = \kappa_i(\boldsymbol{\beta})$ and $Var(C_i|\mathbf{x}_i) = \kappa_i(\boldsymbol{\beta})(1+\alpha\kappa_i(\boldsymbol{\beta}))$. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean, thereby accommodating over or underdispersion. In contrast a semiparametric estimator takes the conditional mean $exp(\mathbf{x}_i'\boldsymbol{\beta}+\omega_i)$ and makes minimal assumptions as to the density of ω_i, other than ω_i is a sequence of i.i.d random variables with mean zero and uncorrelated with $\mathbf{x}_i$.[17]

A flexible *partially* parametric specification can also be achieved by modelling the relationship between the mean and variance, with the variance not fully specified. Estimation is then based upon the Poisson pseudo maximum likelihood estimator, $\widehat{\boldsymbol{\beta}}_{PMLE}$, which has the same first-order conditions as the baseline Poisson model and is consistent for $\boldsymbol{\beta}$ under the weaker assumption of correct specification of the conditional mean. The covariance matrix for $\widehat{\boldsymbol{\beta}}_{PMLE}$ is given by

$$\mathrm{Var}[\widehat{\boldsymbol{\beta}}_{\mathrm{PMLE}}] = \left(\sum_{i=1}^n \kappa_i(\boldsymbol{\beta})\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^n g_i\mathbf{x}_i\mathbf{x}_i'\right)\left(\sum_{i=1}^n \kappa_i(\boldsymbol{\beta})\mathbf{x}_i\mathbf{x}_i'\right)^{-1}, \tag{14}$$

[16]This ensures a closed form expression for (13).

[17]See Zhen (2008).

where $g_i = \text{Var}[C_i|\mathbf{x}_i]$ is the conditional variance of C_i. In the case where $g_i = \kappa_i(\boldsymbol{\beta})$ we have the Poisson model and

$$\text{Var}[\widehat{\boldsymbol{\beta}}_{\text{P}}] = \left(\sum_{i=1}^{n} \kappa_i(\boldsymbol{\beta})\mathbf{x}_i\mathbf{x}_i'\right)^{-1}. \tag{15}$$

For $Var[C_i|\mathbf{x}_i] = (1+\alpha\kappa_i(\boldsymbol{\beta}))\kappa_i(\boldsymbol{\beta})$ and $Var[C_i|\mathbf{x}_i] = (1+\alpha)\kappa_i(\boldsymbol{\beta})$, we have variants of the negative binomial model, respectively NB2 and NB1. For $\text{Var}[C_i|\mathbf{x}_i] = E[(C_i - \mathbf{x}_i'\boldsymbol{\beta})^2|\mathbf{x}_i]$ unspecified a consistent estimate of $\text{Var}[\widehat{\boldsymbol{\beta}}_{PMLE}]$ is a variant of the *Eicker-White* robust sandwich variance estimator. As noted by Cameron and Trivedi (2005) the modification of the Poisson regression model in this way represents an example of the generalised linear model (GLM) (see Nelder and Wedderburn (1972)). The essential characteristic of these models is that consistency requires a correct specification of the conditional mean.

In this paper we adopt an alternative flexible estimation framework, namely the Generalised Method of Moments (GMM). Given that the mean of a count process is non-negative then it is convenient to write the conditional mean of CEB as an exponential function of a linear index in the regressors. A GMM approach is particularly suited to the estimation of the parameters of this non-linear regression function since by using first-order conditions only it allows us to depart from a highly restrictive Poisson specification, allow for unobserved heterogeneity without a full set of parametric assumptions, and accommodate endogenous regressors. The GMM estimator is based on the moment conditions $E(v_i|\mathbf{x}_i) = 1$, where $v_i = \exp(\omega_i)$. To see this note that the conditional mean function with multiplicative errors is given by

$$E(C_i|\mathbf{x}_i, \omega_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \omega_i) = \kappa_i(\boldsymbol{\beta})v_i. \tag{16}$$

The conditional expectation of C_i with respect to $\mathbf{x}_i$ is then given by $E(C_i|\mathbf{x}_i) = \kappa_i(\boldsymbol{\beta})E(v_i|\mathbf{x}_i)$.[18] In the face of endogenous regressors such that $E[v_i|\mathbf{x}_i] \neq 1$, the conditional moment restrictions are given by $E(v_i - 1|\mathbf{z}_i) = 0$. The GMM estimator minimises the (*generalized*) length of the empirical vector function

$$\hat{\boldsymbol{\beta}}_{GMM} \equiv \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}} \boldsymbol{\psi}_N(\mathbf{w}, \boldsymbol{\beta})'\tilde{Q}_N^{-1}\boldsymbol{\psi}_N(\mathbf{w}, \boldsymbol{\beta}), \tag{17}$$

where $\mathbf{w}$ collects all observables including C, potentially endogenous regressors $\mathbf{x}$, and instrumental variables $\mathbf{z}$. Given a multiplicative error model then $\boldsymbol{\psi}_N(\mathbf{w}, \boldsymbol{\beta}) = \mathbf{Z}'(\mathbf{y} -$

[18]We note that given the use of a multiplicative model then any additional underlying regression error would be absorbed by the unobserved heterogeneity term ω_i. See Mullahy (1997) for an extended discussion of this point.

$\kappa(\boldsymbol{\beta})/\kappa(\boldsymbol{\beta}))$ such that we can write (17) as

$$\hat{\boldsymbol{\beta}}_{GMM} \equiv \arg\min_{\boldsymbol{\beta}\in\boldsymbol{\Theta}}[\frac{1}{n}\mathbf{Z}'(\mathbf{C}-\kappa(\boldsymbol{\beta})/\kappa(\boldsymbol{\beta}))]'\tilde{Q}_N^{-1}[\frac{1}{n}\mathbf{Z}'(\mathbf{y}-\kappa(\boldsymbol{\beta})/\kappa(\boldsymbol{\beta}))],$$

where $\kappa(\boldsymbol{\beta}) = \exp(\mathbf{X}'\boldsymbol{\beta})$ and $\mathbf{Z} = \{\mathbf{z}_i\}$ denotes an instrument matrix.[19] $\tilde{Q}_N = \mathbf{Z}'\Omega\mathbf{Z}$ denotes the optimal weight matrix where $\Omega = (C - \kappa(\widehat{\boldsymbol{\beta}})/\kappa(\widehat{\boldsymbol{\beta}}))^2$. $\kappa(\widehat{\boldsymbol{\beta}}) = \exp(\mathbf{X}'\widehat{\boldsymbol{\beta}})$ with $\widehat{\boldsymbol{\beta}}$ a first step GMM estimator based upon any weighting matrix.

6 Conditional Mean Specification

In this study we extend the Beckerian utility model for children ever born to account for the impact of interactions across households which belong to distinct ethnic groups and reside in clusters of households.[20] We write the conditional mean as

$$E(C|\mathbf{x}^I, \mathbf{h}, \mathbf{p}, \mathfrak{e}) = \exp(\mathfrak{e} + \beta^c E[f(C|\mathfrak{cl})] + \boldsymbol{\gamma}\overline{ed}_{cl} + \mathbf{h}'\boldsymbol{\theta} + \mathbf{p}'\boldsymbol{\kappa} + \mathbf{x}^{I\prime}\boldsymbol{\eta}). \tag{18}$$

(18) includes the standard set of individual ($\mathbf{x}^I$), household ($\mathbf{h}$), and cluster ($\mathbf{p}$) characteristics, with associated vectors of parameters $\boldsymbol{\eta}$, $\boldsymbol{\theta}$, and $\boldsymbol{\kappa}$. β^c and $\boldsymbol{\gamma}$ are scalar parameters denoting, respectively, endogenous and exogenous effects.[21] $\mathfrak{e}$ denotes a categorical variable denoting ethnic affiliation.

Below we turn our attention to the specification of the endogenous and exogenous effects, integrating the econometric specification with the theoretical model outlined in section 4.

6.1 Endogenous and Exogenous Interaction Effects

In section 4 we discussed a model in which endogenous local neighbourhood effects and exogenous global ethnic effects might have influence on fertility decisions. We represent endogenous effects attributable to the fertility choices of other women in the same cluster using

$$\beta^c E[f(C|\mathfrak{cl})] = \partial S_2^{\mathfrak{e}}(C_i, \overline{C}_{-i}^{\mathfrak{cl}})/\partial C_i = \beta^c(\overline{C}_{\mathfrak{cl}-i} - C_i), \tag{19}$$

[19] The GMM estimator $\boldsymbol{\beta}_{GMM}$ accounts for both the endogeneity of one or more of the elements of x and the intrinsic heteroscedasticity of $\kappa_i(\boldsymbol{\beta}) = \exp(\mathbf{x}_i'\boldsymbol{\beta})$.

[20] As such the fundamental characteristic which differentiates our approach from multilevel (variance components) techniques is that we assign, theoretically as well as empirically, a more prominent role to the mechanisms by which social interactions affect fertility behaviour. Although the failure to account for dependencies across observations can result in inefficient estimation, our focus is to build an appropriate structural model of the interactions. See Moulton (1986), Chamberlain (1980) and Mundlak (1978).

[21] For clarity of exposition we restrict β^c to be a scalar quantity. However, in the empirical section we allow endogenous effects to vary across ethnic groups such that β^c is now a vector.

where C_i denotes the number of children ever born to woman i, and $\overline{C}_{\mathfrak{cl}-i}$ is the mean CEB for cluster $\mathfrak{cl}$. β^c represents the cross-partial effect[22] $\frac{\partial^2 S_2^{\mathfrak{c}}(.)}{\partial C_i \partial \overline{C}_{-i}^{\mathfrak{cl}}}$.

We represent exogenous effects attributable to *observing* the education of other women in the same cluster using $\overline{ed}_{cl} = \frac{1}{n_{cl}} \sum\limits_{i \in cl} ed_i$ with ed_i denoting the education level of the i^{th} woman residing in the same cluster, and n_{cl} denotes the number of individuals resident in the cluster. It is argued that the effect of education is iconic, with some individuals aspiring to the attributes of other higher educated members within their cluster. We choose to model these education interactions at the level of the cluster as higher educated women in the locality may be leaders and opinion-makers who influence others who are less educated to adopt low fertility norms. Hence we consider the proportion of women in the cluster who had completed higher education as representative of this exogenous education effect on fertility.[23]

We also include a number of other cluster *attributes* to reduce the likelihood of spurious inference. Critical in this respect is access to water and fuel infrastructure. Based upon the work of Dasgupta (2000), better access to water and fuel reduces the demand for child labour to collect them, and that this is turn reduces the demand for children, and hence the fertility rate (Iyer (2002)). Access to fuel was measured by a mean-level effect for access to electricity. Access to water infrastructure at the cluster-level was measured according to whether access to water was located (i) in the residence (either piped into the residence or a well was located in the residence); (ii) obtained from a publicly-provided source (either a public tap or a public well); (iii) obtained by collecting it from a river, stream, pond or lake; or (iv) whether rainwater was relied upon as the chief source of water. In addition, an additional cluster-level attribute included was whether or not a radio was listened to at least once a week which again was aggregated up from whether or not an individual listened to a radio.

6.2 Household-level and Individual-level Controls

We control for a number of household and individual characteristics that we know might also have an impact on fertility decisions to ensure that the endogenous and exogenous effects observed at the cluster level are robust to variations in other characteristics. The choice of variables is in keeping with a long lineage of economic models of fertility decisions. These factors include household structure (whether it is polygytnous or not); the

[22]The notation $S_1()$ and $S_2()$ refers, respectively, to the proportionate and conformist social utility functions presented in Section 4.

[23]It is important to clarify that we are using a proportionate metric for the exogenous effect as we believe that the average level of education has an influence on fertility behaviour due to the 'iconic' value of education discussed earlier, rather than the distance of a particular woman's educational attainment from the attainment of better-educated women in the group. Therefore we argue that the proportionate metric is a better representation of the woman's actual behaviour than the conformist metric, at least with reference to the effect of education on fertility outcomes.

number of other household residents; the age of the woman; household income and other characteristics.

Polygyny can have two impacts on fertility: first, it can increase fertility if there is more help provided with child-care, creating a micro-level externality within the extended family household. Second, if there is greater discussion of family planning issues, then this might work to reduce fertility. Alternatively if high-fertility norms are espoused within the extended household then the effect might be to increase fertility.[24] There are 30% of households in the sample that had more than one wife living in the household and in our model we control for the number of usual residents in the household. The rationale for the inclusion of this variate is that if other household members help with child care, or indeed as with other African societies where 'fosterage' is common, other residents have an important bearing on a woman's total fertility.[25]

We also include controls for income and other characteristics. One notable problem here is that in the DHS no direct questions on income were asked of survey respondents. Utilising a series of questions asked about household quality, access to infrastructure, and the ownership of consumer durables, it was possible to construct a number of indicators which control for the economic status of households. The quality of roof construction was measured on the following (increasing quality) scale: iron (mabati), tiles, grass or thatch, and other material. The quality of the floor was measured using (decreasing quality) scale: mud, dung or sand, wood planks, tiles or polished wood, and cement. Other indicators of the quality of housing infrastructure were whether or not it had a toilet, and the number of rooms for sleeping. The status of the household was also proxied by ownership of consumer durables including whether or not the household owns a radio, a television, a telephone, and a bicycle. We also include a mortality variable here that measures whether the woman had a son or daughter who had died.

Two other important household-level attributes that we also use are (i) a measure of infant mortality, and (ii) reported awareness of HIV/AIDS in the population. These factors are included in order to control for the fact that Kenya, as like other countries in Africa, has seen a large increase in HIV-related deaths over the time period under study which might also affect fertility. This is because demographers think that higher infant and child mortality is frequently associated with higher fertility rates if parents want to replace children who are lost to child death, or if they perceive that the average mortality rate in their region is high. So we control for mortality changes which might also influence fertility in this country.

Individual-level controls include the age of the woman measured in years and an age-

[24] A number of recent studies have examined the role of polygynous marriage norms in influencing demographic behaviour. See for example Hogan and Biratu (2004).

[25] A number of other demographic studies of poor societies (see, for example, Iyer (2002)) have found that the role played by residents within the household or within close proximity, such as friends and neighbours, is important for fertility.

squared variable accounting for the non-linearity associated with age-related variables. The education of the woman is included and measured as the highest level of education attained, separating out primary, secondary, and higher education effects. The influence of the media was measured as whether the woman listened to the radio at least once a week. This variable was included on the assumption that greater information about contraceptive technology would be available to the woman if she listened to the radio.[26] A binary variable recorded how long the woman had lived in the community.[27] This variable was included on the assumption that if a woman has resided in the community for a longer period, then she was more likely to have formed stronger networks in the cluster, which in turn may influence her fertility behaviour.We also include a variable here for whether the individual knows a place to test for HIV/AIDS.

Given that our sample includes women between 15 and 49 years of age, we account for the fact that for younger women we effectively observe truncated fertility lifetimes by adding the variables 'age' and 'age squared' to control for the fact that older women would have completed their childbearing, and therefore might display higher fertility, relative to younger women in the sample. This also addresses the time inconsistency problem in using DHS data that women may be making decisions with respect to their fertility considerably before the time when they were selected for survey and interview, and that therefore their current economic status (as for example measured by ownership of consumer durables, or the current state of the roof quality or floor quality in their homes) may not be as representative of their economic status at the time that they made their decision about having a child, as has been pointed out by Arulampalam and Bhalotra (2003).

7 The Identification Problem

The nature of the identification problem in models of social interaction depends upon the observed data, the model, and whether the analyst seeks to identify exogenous and endogenous effects. For example, if we rewrite the conditional mean from (18) within a linear framework, and allow for the existence of exogenous effects for *all* individual characteristics, namely

$$E(C|\mathbf{x}^I, \mathbf{h}, \mathbf{p}, \mathfrak{c}) = \mathfrak{c} + \beta^c E(\overline{C}_{\mathfrak{cl}-i} - C_i|\mathfrak{cl}) + \boldsymbol{\gamma} E[\mathbf{x}^I|\mathfrak{cl}] + \mathbf{h}'\boldsymbol{\theta} + \mathbf{p}'\boldsymbol{\kappa} + \mathbf{x}^{I\prime}\boldsymbol{\eta}, \qquad (20)$$

[26] Messages about this are routinely broadcast on the national radio channel, the 'Voice of Kenya' in English, Swahili and the vernaculars.

[27] If she had lived in the same community since 1993 then this variable took the value 0; if she had migrated in the interim period and had therefore lived in at least two communities, then this variable took the value 1.

then the genesis of the identification problem derives from the simple observation that in (20) we specify a model for children ever born which is linear in a mean endogenous effect, $E(\overline{C}_{\mathfrak{cl}-i} - C_i|\mathfrak{cl})$, and a vector of mean exogenous effects, $E[\mathbf{x}^I|\mathfrak{cl}]$. It therefore follows that the reduced form representation of the conditional mean $E(C|\mathbf{x}^I, \mathbf{h}, \mathbf{p}, \mathfrak{e})$ will be a function of $E(\mathbf{x}|\mathfrak{cl})$.[28] This is obviously central to the question of identification. Namely we start with the intention of disentangling a mean group (cluster) effect, in terms of the three components $E(\overline{C}_{\mathfrak{cl}-i} - C_i|\mathfrak{cl})$, $E(\mathbf{x}^I|\mathfrak{cl})$, and $E(u|\mathbf{x}^I, \mathbf{h}, \mathbf{p}, \mathfrak{cl})$, where u is the error in the underlying linear regression model. We then observe that, not surprisingly, the reduced form $E[\overline{C}_{\mathfrak{cl}-i} - C_i|\mathfrak{cl})]$ is a function of $E(\mathbf{x}|\mathfrak{cl})$.

There are a number of ways to circumvent the identification problem.[29] Within a linear framework the simplest yet probably the most restrictive method, is to impose a full set of zero restrictions on $\boldsymbol{\gamma} = 0$, thereby adopting the assumption that any interactions are mediated solely by endogenous effects. We note that such an approach is potentially problematic for making inference on multiplier effects, since such restrictions may result in biased estimates of pure endogenous effects. This is important since policy proscriptions which follow from endogenous versus exogenous effects are likely to be very different. However, the upside of such an approach is that these restrictions automatically generate a possible set of instruments $(E(\mathbf{x}|\mathfrak{cl}))$.[30] Second, and as discussed earlier, although it is possible to conceive of a large number of exogenous effects, in the specific context of fertility behaviour education plays a particularly prominent role. As a result and following Brock and Durlauf (2000), identification is facilitated by locating individual level variables whose group level average does not enter the conditional mean.

In addition to the imposition of a partial set of zero restrictions on $\boldsymbol{\gamma}$, identification is achieved as a consequence of a correctly specified functional form. Namely, we apply a link function to the conditional expectation $E(C|\mathbf{x}^I, \mathbf{h}, \mathbf{p}, \mathfrak{e})$ so as to correctly represent the nature of the observed data. We note that identification is then achieved as a result of the specified nonlinear model, rather than as a specifically targeted means to blur the linear dependence between endogenous and exogenous effects. The conditional mean is then written as (18). As a consequence the linear dependence between these two objects is automatically removed given that the we have imposed a restriction on the support of $E(C|\mathfrak{cl})$: between zero and $\varkappa$, where $\varkappa$ is the maximum number of children ever born.[31]

[28] See Manski (1993) for further details.

[29] Graham (2008).

[30] See, for example, Case and Katz (1991) and Gaviria and Raphael (2001). See Kawaguchi (2004) for an approach which utilises subjective perceptions of peers' average behaviour, instead of average behaviour.

[31] See Brock and Durlauf (2003) for a discussion of multinomial choice with social interactions.

7.1 Endogeneity Bias and the Use of Metereological Data as an Instrument for Fertility

In specifying the conditional mean, endogenous cluster conformist effects are given by $(\overline{C}_{\mathfrak{cl}_{-i}} - C_i)$, for C_i denoting children ever born to woman i, and $\overline{C}_{\mathfrak{cl}-i}$ denoting mean fertility in the cluster. To circumvent the problem of endogeneity we make the case that observed variations in actual fertility rates across regions reflect the time-series and geographical variations in rainfall. We contend that for tropical countries like Kenya which both experience hot weather, and exhibit less fertility control than in developed countries, properties of the distribution of rainfall, and in particular the mean and variance, represent a valid instrument for the observed variation in fertility. Given that ethnic groups tend to reside in distinct regions then geographic variation in rainfall is likely to explain variation in fertility rates by ethnicity.

The use of rainfall data as an instrument is based on a number of arguments. First, in low-income countries in which fertility regulation is not practiced extensively, and in which parents calculate the costs and benefits of having an additional child based on factors like their income (Becker (1981)), rainfall variations will influence fertility behaviour. For example, in periods of drought when income is likely to be low, parents might postpone childbearing (Lam and Miron (1996)). Second, high temperatures will have an effect on conception and thereby on fertility. Combining these two economic and biological arguments, we note that in areas of abundant rainfall, which are on average cooler and have lower temperatures, we expect *a priori* that these areas will also be areas of higher fertility. This is because in areas of abundant rainfall, income will be high. Areas that are very arid and receive on average very poor rainfall are also likely to be those with very high hot weather temperatures, and will exhibit lower fertility as conception rates are likely to be low. We would therefore expect to observe higher fertility in better rainfall-fed districts of Kenya as compared to those which have lower rainfall.

In addition to the impact of mean rainfall, we argue that the seasonality of rainfall exerts an additional independent effect on fertility decisions, and as such represents an additional instrument. The argument for the use of this instrument is based on the notion that seasonality of rainfall proxies uncertainty.[32] The impact of a reduction in uncertainty on fertility can be either positive or negative. If parents view children as an 'investment' good then a decrease in uncertainty can reduce the insurance value of having children and thus decrease the net benefit of having children (Iyer and Velu (2006)). Alternately, if children or 'child services' are viewed purely as a 'consumption' good, then a decrease in uncertainty is likely to result in parents that are more likely to have more children as the desired number of children rises. The net impact on fertility will depend upon the

[32] Recent papers in economic demography have also been examining the impact of uncertainty and the real options approach in order to understand decision-making about fertility (see, for example, Iyer and Velu (2006); Bhaumik and Nugent (2008).

relative effects of the decrease in uncertainty and the decreased benefit of the insurance effect of children (see Iyer and Velu (2006) for a thorough discussion).

8 Data and characteristics of DHS survey

Reflecting the population breakdown in the country as a whole, our sample, taken from the Kenyan Demographic and Health Survey (KDHS)Ken (1999),[33] has data on the Kalenjin, the Kikuyu, Luhya, Luo, Kamba, Kisii, Mijikenda/Swahili and Meru/Embu.[34] The Kenya DHS adopted a two-stage stratified sampling approach that selected households located within primary sampling units (PSU) or sampling clusters. These clusters are identical to the complete enumeration of sample clusters which took place as part of the 1977 National Demographic Survey. The sample points themselves are identical to those chosen in the sampling frame maintained by the Kenyan Central Bureau of Statistics. In the 35 of Kenya's 42 districts that were included in the survey, there were 536 clusters - 444 rural clusters and 92 urban clusters, of which 530 were non-empty clusters.[35] The location of the clusters geographically are identifiable within the district and province (DHS and Macro International, 1999: 179-182). A complete list of all households in each cluster was recorded between November 1997 and February 1998. From the remaining 530 clusters, a systematic sample was drawn of, on average, 22 households in urban clusters and 17 households in rural clusters. This formed a total of 9465 households. In these households, all women age 15-49 were targeted for interview. Response rates varied by province from approximately 88% to 99% (for more details see KDHS, 1998 p.180).

In addition to the geographical stratification of data by cluster and household, it is also possible to stratify these data by ethnicity. As also discussed in section 2, the major Kenyan ethnic groups (covering 88% of the population) are Kikuyu, Luo, Luhya, Kamba, Kalenjin, Kisii and Mijikenda/Swahili. The Kikuyu is the largest, with 17.9% of the population, while Mijikenda/Swahili is the smallest, with 5.0%. The representation of each group in the sample is similar to its representation in the whole population. The largest number of women included in the sample live in the Rift Valley region, while the smallest number of women sampled live in Nairobi. Looking at the distribution by ethnic group, the largest sample was drawn from the Kalenjin group, while the smallest was

[33] The survey contains interview data for 7800 women aged 15-49. It contains women from all of Kenya's large ethnic groups and covers a wide geographical area, omitting only the areas of extremely low population density in the North. Geographically. Kenya is divided into 7 provinces which are further subdivided into 47 districts. The Kenya DHS covered 42 of Kenya's districts; 35 were sampled, 7 were not. 17 districts were oversampled.

[34] Our study excludes the Maasai, a pastoral group living in the Kenyan Rift Valley, and other tribes such as the Galla and Somalis, who live in the north-east.

[35] 6 of the clusters could not be included in the survey due to inaccessibility (DHS and Macro International 1999: 180).

drawn from the Meru/Embu group.

According to the DHS data, a rural woman has on average about 5.2 children compared to fertility among urban women at 3.1 children. Fertility differentials by the level of education show that illiterate women bear on average 5.8 children compared to 3.5 children for women with secondary school education (DHS 1998, p. xvii). The sample reveals demographic differences between ethnic groups: for women aged 40-49, the mean CEB varies from 5.91 for Kikuyu to 7.56 for Luo. The differences in fertility by ethnic group are clearly very large. Table 1 shows the mean CEB grouped by region in the sample. CEB varies from a low of 1.7 in Nairobi/Central to a high of 3.3 in Nyanza. This pattern is also seen by region. The Kikuyu, Kalenjin, Luhya and the Luo follow similar patterns. For example, CEB is low among the Kikuyu (in the Nairobi region), and high amongst the Luo (in the Nyanza region).

The DHS data show that knowledge of family planning in Kenya is very high: 98% of women and 99% of men were able to name at least one modern method of contraception. There are 39% of women who use contraception and the most widely used methods are contraceptive injectables, the pill, female sterilization and periodic abstinence. Contraceptive use does however vary greatly by region: while there are 61% of women in the Central province who use contraception, only 22% of women in the Coast province do so likewise. Only 23% of women with no education use contraception compared to 57% of women with secondary education. Both government and private medical sources provide access to contraceptives.

In the data used in the present study we consider individual-level data only for the 5994 women residing in a rural cluster, and who did respond. Table 3 presents the counts of children ever born in the sample.

8.1 Constructing the rainfall instruments

In this study we use unique historical meteorological data on monthly rainfall collected for us by the Kenyan Meteorological Department from thirty-one meteorological stations in Kenya at 5-year intervals, from 1930 to 2005. This data was aggregated into yearly average rainfall and its standard deviation by rainfall station in each Kenyan district. Below we outline the matching exercise that allowed us to assign to each woman for each birth that she underwent, an average annual rainfall estimate for that district at the time closest to the year of the birth of her child.

The matching of the metereological and the DHS data proceeded by assigning, for each observation in a particular district, rainfall data which minimised the Euclidean distance between the district capital and the nearest metereological station.[36] To quantify the impact of uncertainty on fertility, we construct an index for the seasonality of the rainfall

[36] Using latitude and longitude data for the district capitals and those of the meteorological stations.

across the meteorological stations in Kenya over the course of a year. For each year the index takes the difference between the average rainfall in the wettest and the driest three months, and scales them by the average rainfall for the year for that particular meteorological station (see Foeken (1994)). This is reported as the index of District Rainfall Seasonality (DRS). We also report the standard deviation of mean rainfall in 2005.

Table 4 depicts the geographical variations in rainfall across the different provinces of Kenya. Rainfall is highest on the western coast of Kenya (the Western province and Nyanza) in which fertility rates are also the highest (3.2 and 3.3). Mean rainfall is lowest in Nairobi which also depicts the lowest fertility (1.6). As shown in Table 5, rainfall in Kenya is highly seasonal and varies considerably by geographical region. We examine correlations between the total number of children ever born and the average rainfall by district in Kenya. This correlation is positive and is 0.437 for our measure of total fertility children ever born with average rainfall by district. For women who have 10 births or fewer, this correlation was even stronger at 0.4598. Table 6 shows for the year 1998 children ever born by province, mean rainfall and rainfall seasonality for the different Kenyan regions.

Looking at Tables 4, 5 and 6 collectively, suggests that lower rainfall seasonality might be associated with higher fertility: the greatest seasonality of rainfall is in Nairobi, the Central province and the eastern province, the three regions that also have the lowest fertility. Of course these findings could also be driven by other characteristics of these provinces such as income or education, and in the models presented in Section 10 below we examine the effect of our rainfall instrument after controlling for some of these other factors. Suffice to say at this stage however that if we look simply at the correlations between these rainfall patterns and fertility rates across the different provinces of Kenya, it would appear that the areas with the *highest rainfall, and lowest rainfall seasonality, are also those in Kenya which show the highest fertility.*

We therefore contend that by using rainfall as an instrument both with respect to its absolute amounts as measured independently at metereological stations, and by considering also its seasonality that differs across the different provinces of Kenya, this measure provides independent exogenous variation for the endogenous fertility behaviour.

9 Results

Table 7 presents our results using the GMM estimator with the number of children ever born (CEB) instrumented using mean rainfall and the seasonality index. The dependent variable in the model is the number of children ever born. Parameter estimates may be interpreted as the proportional change in the number of children ever born due to a unit change in the regressors.[37] Since the predicted count is given by $e^{\mathbf{x}'\hat{\boldsymbol{\beta}}}$, we also report the factor change in the expected count for a unit increase in a given x_k, denoted $e^{\hat{\boldsymbol{\beta}}}$, and emphasise this statistic when interpreting our results. We report bootstrap standard errors using 150 replications and use the Huber (1967) method to account for any residual intra-cluster correlation.

Endogenous effects The localised endogenous effects emanating from within the village cluster, are captured by a parameter which, for each ethnic group (e), weights the variable ($\overline{C}_{e,cl} - C_i$) - namely the difference between a mean-level and individual CEB (instrumented by rainfall) for women who live in the same cluster. We observe that for 7 out of 8 groups these parameter are significant and positive with effects ranging from 1.189 for the Kikuyu, 1.20 for the Kamba, 1.17 for the Luhya, to 1.11 for the Luo. The only group for whom these effects were not significant was the Kisii. A key aspect of these results is that consistent with our anthropological understanding of the ethnic groups (and as discussed in section 3), for those clusters in which individuals were related both by blood and marriage, as for example, among the Kikuyu *mbari*, the endogenous effects on fertility are among the stronger of the ethnic group effects on fertility. We observe that those groups who are related both by blood and by marriage, as for example the Kikuyu, exhibit significant and large interaction effects. The importance of these effects is that they suggest more clearly the channels through which the effect of ethnicity impacts on fertility. This is important for demographic analyses per se since our results in indicate the importance for fertility of including a measure which captures the effects of interactions within ethnic groups, and the fact that they are important over and above the significance of individual and household characteristics.

Exogenous effects; cluster-level controls We consider exogenous effects as attributable to the effect of observing the level of education of other women in the same cluster. We measure this effect using the proportion of women who had completed higher education, allowing parameters to differ by ethnicity and the age of the woman[38]. The effects in Table 7 are insignificant after controlling for the individual levels of educational attainment for all ethnic groups, except the Luo for whom this variable in significant at the 10% level and for whom the parameter estimate was of the order of 0.149. In order to reduce the likelihood of omitted cluster attributes confounding our inferences on exogenous and endogenous interactions a range of cluster-level controls are included: these are media access and access to water and fuel infrastructure in the cluster. Among

[37] This follows since

$$\frac{\partial E(C|\mathbf{x})}{\partial x_j} = \beta_j \exp(\mathbf{x}'\boldsymbol{\beta})$$

where $E(C|\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta})$.

[38] For the Mijikanda-Swahili and Kamba ethnic group the numbers of women who had completed higher education was extremely small across most clusters, and as a result, the exogenous education effect was constructed utilising the proportion who had completed a secondary school education.

these cluster-level controls, the effect of the media was significant and exerted a negative effect on fertility, with a factor effect of 0.886 decrease. None of the mean-level cluster effects for water or electricity were significant.

Household-level effects; regions and ethnic controls At the household level we make a distinction between interactions within the household which may have implications for fertility, and household-level controls. Household-level controls are included so as to reduce the possibility that any inference on household-level interaction effects is not confounded by household-level omitted variables. None of the household-level social interaction effect were significant nor was the infant mortality variable at the household level. The infant mortality variable measured whether a son or a daughter had died.

In general we would expect that income would be negatively correlated with fertility. As noted above, one of the problems with the DHS survey is that no direct questions on income were asked of survey respondents. To address this deficiency a number of household level variables were included as controls for income and other characteristics. Measures of roofing quality are included to suggest that if the woman lived in a household that was wealthier, as measured by a better quality roof such as iron or tiles, then her CEB was more likely to be lower, relatively to the base category - a thatched roof. If a woman lived in a house which had an iron roof, the expected number of children ever born was likely to be lower than if she was resident in a house with a thatched roof. None of these variables emerged significant however. This was also true of the floor quality variables: if the woman lived in a house which had a cement floor (relative to the base of sand), then this might have been important but again this was not significant. Two variables included in the model which could be construed both as income indicators and as controls for media access are radio and television ownership but these were also not significant at the household level.

All the ethnicity effects were significant: Kalenjin (1.244), Luhya (1.196), Mijikenda-Swahili (1.151) and the Luo (1.410) who all showed much higher fertility than the Kikuyu who formed the base group. Given that in many cases ethnic groups settle in specific regions, for example 74 % of the population of coastal province are Mijikenda-Swahili, it was not possible to separately identify all regional effects. Of those regional effects that we could identify none were significant.

Individual-level effects As expected many of the individual-level variables were highly significant: these included the woman's age with an older woman exhibiting higher fertility compared to her younger counterpart. This is something we might expect if norms concerning family size have evolved over time. Women today are more likely on average to have fewer numbers of children compared to previous generations of women before them. For women, primary and higher education were significant, and with the expected signs of positive for primary education and strongly negative for higher education. The effects of higher education reduces fertility by, on average 0.851. The base category was women with no education. Primary education has the expected positive relationship with CEB.

We also note that if the woman had lived continuously in the community since 1993, then this was likely to increase her CEB by about 0.900. One reason for this finding may be that women who stayed more continuously in one region were more likely to benefit from the impact of the greater 'social capital' generated by living in the community for long periods. It is also possible that they are more likely to form stronger networks with others if they have lived there longer, and this would be particularly important for fertility-related issues. In contrast, another factor that was not very important at the individual level was whether the woman had an awareness of a place to go to test for HIV/AIDS (measured by a variable KnowPlaceAids).

10 Conclusion

Strategic complementarities in fertility decisions implies that a couple's fertility decisions may be dependent on the actions of others in the vicinity or in the society more widely. The mechanisms through which these complementarities occur are through social interactions. This paper has examined the importance of social interactions in the context of fertility behaviour in Kenya. We have examined the role of dependencies across individuals that reside in a cluster of households who locate according to ethnicity. By identifying the multiplicity of these channels, we have a better basis upon which to attempt to influence policy. More significantly, the existence of multiple channels of social interaction imply that an analyst attempting to isolate these pathways, needs to be cautious about the possibility of erroneous inference.

Collectively, our results enable us to determine the relative importance of individual determinants of fertility alongside a number of channels by which fertility decisions are influenced by the behaviour of others. We have demonstrated the significance of conformist endogenous effects on fertility operating at the level of the village cluster, and differentiated by ethnicity. After controlling for individual education, exogenous effects were generally not a significant determinant of fertility. Similarly our results suggest that, after controlling for a large number of individual effects and other forms of interaction, the household level interactions are not important. However, our results have confirmed strongly the importance of ethnic-level effects, alongside the traditional determinants of fertility at the individual level. Collectively, therefore we argue that social interaction effects matter greatly for fertility decisions in Kenya.

References

(1970): *Population Census 1969* Republic of Kenya, Vol. 1, Nairobi.

(1999): *Kenya: Demographic and Health Survey*National Council for Population and Development (NCPD), Central Bureau of Statistics (CBS), Office of the Vice President and Minsitry of Planning and National Development (Kenya) and Macro International Inc. (MI), Calverton, Maryland.

Ajayi, A., and J. Kekovole (1998): "Kenya's Population Policy from Apathy to Effectiveness," in *Do Population Policies Matter? Fertility and Politics in Egypt, India, Kenya and Mexico*, ed. by A. Jain. New York: Population Council.

Akerlof, G. (1997): "Social Distance and Social Decisions," *Econometrica*, 65, 1005–1027.

Arulampalam, W., and S. Bhalotra (2003): "Sibling Death Clustering in India: Genuine Scarring vs Unobserved Heterogeneity," Department of Economics, University of Bristol, UK, Bristol Economics Discussion Papers 2003.

Ascadi, T. F., G. J. Ascadi, and R. A. Bulatao (1990): "The Context and Causes of High Fertility," in *Population Growth and Reproduction in Sub-Saharan Africa: Technical Analysis of Fertility and its Consequences*, ed. by T. F. Ascadi, G. J. Ascadi, and R. A. Bulatao, pp. 131–132, Washington, D.C. The World Bank.

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

Becker, G. S., and K. M. Murphy (2000): *Social Economics: Market Behavior in a Social Environment.* The Belknap Press of Harvard University Press, Cambridge, M.A. USA.

Berg-Schlosser, D. (1984): "Tradition and Change in Kenya," *International Gegenwart Band 3*, Paderborn: Ferninand Schoningh.

Borjas, G. J. (1995): "Ethnicity, Neighbourhoods and Human-Capital Externalities," *American Economic Review*, 85(3), 365–390.

Brock, W., and S. Durlauf (2000): "Interactions Based Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Learner, no. 5. North Holland Press, Amsterdam.

Brock, W. A., and S. N. Durlauf (2001): "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 235–260.

Brock, W. A., and S. N. Durlauf (2003): "Multinomial Choice with Social Interactions," National Bureau of Economic Research, Inc, NBER Technical Working Paper: 288.

Caldwell, C., and P. Caldwell (1987): "The Cultural Context of High Fertility in Sub-Saharan Africa," *Population and Development Review*, 13(3), 409–437.

Case, Anne, C., and F. Katz, Lawrence (1991): "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," National Bureau of Economic Research, Inc, NBER Working Papers: 3705.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225–228.

Coale, A. J., and S. C. Watkins (eds.) (1986): *The Decline of Fertility in Europe*Princeton. Princeton University Press.

Conley, T., and G. Topa (2002): "Socio-Economic Distance and Spatial Patterns of Unemployment," *Journal of Applied Econometrics*, 17, 353–395.

Cooper, J. (1988): "Coordinating Coordination Failures in Keynesian Models," *The Quarterly Journal of Economics*, 103(3), 441–463.

Dasgupta, P. (2000): "Population and Resources: An Exploration of Reproductive and Environmental Externalities," *Population and development review*, 26, 643–689.

De Wilde, J. C. (1967): *Experiences with Agricultural Development in Tropical Africa*, vol. II: The Case Studies. John Hopkins Press, Baltimore.

Durlauf, S. N., and J. Walker (1999): "Social Interactions and Fertility Transitions," Paper prepared for workshop on 'Social processes underlying fertility change in developing countries.', 28-30 Jan 1998, organised by the Committee on Population, National Academy of Sciences.

Fapohunda, B. M., and P. V. Poukouta (1997): "Trends and Differentials in Desired Family Size in Kenya," *African Population Studies/Etude de la Population Africaine*, 12(1).

Federici, N., K. O. Mason, and S. Sogner (eds.) (1993): *Women's Position and Demographic Change*Oxford. Clarendon Press.

Ferguson, K., and T. SrungBoonmee (2003): "Culture and the Economy in Kenya," Paper submitted for the 1st Annual undergraduate research symposium, University of South Florida, Spring.

Foeken, D. (1994): "Climatic Seasonality in Kenya with Special Reference to Coast Province," *Eastern and Southern Africa Geographical Journal*, 5(1), 28–43.

Gaviria, A., and S. Raphael (2001): "School-Based Peer Effects and Juvenile Behavior," *Review of Economics and Statistics*, 83(2), 257–268.

Ginther, D., R. Haverman, and B. Wolfe (2000): "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust are Relationships?," *The Journal of Human Resources*, XXXV(4).

Glaeser, E. L., and J. A. Scheinkman (1999): "Measuring Social Interactions," Working Paper, Harvard University.

Graham, B. S. (2008): "Identifying Social Interactions Through Conditional Variance Restrictions," *Econometrica*, 76(3), 643–660.

Guinnane, T. W., C. Moehling, and C. O'Grada (2001): "Fertility in South Dublin a Century Ago: A First Look," Yale University Economic Growth Centre Discussion Paper No. 838.

Hank, K. (2001): "Regional Social Contexts and Individual Fertility Decisions: A Multilevel Analysis of First and Second Births in Western Germany," MPIDR Working Paper No. 2001-015.

Hogan, D. P., and B. Biratu (2004): "Social Identity and Community Effects on Contiagitive Use and Intentions in Southern Ethiopia," *Studies in Family Planning*, 35(2), 79–90.

Horst, U., and J. A. Scheinkman (2003a): "Equilibria in Systems of Social Interactions," mimeo Humboldt University of Berlin, Department of Mathematics.

——— (2003b): "Equilibria in Systems of Social Interactions," mimeo, Humboldt University of Berlin, Department of Mathematics.

Iyer, S. (2002): *Demography and Religion in India*. Oxford University Press, Delhi.

Iyer, S. (2002): "Religion and the Decision to Use Contraception in India," *Journal for The Scientific Study of Religion*, 41(4), 711–722.

Iyer, S., and C. Velu (2006): "Real Options and Demographic Decisions," *Journal of Development Economics*, 80(1), 39–58.

Kawaguchi, D. (2004): "Peer Effects on Substance Use Among American Teenagers," *Journal of Population Economics*, 17, 351–367.

Kenyatta, J. (1961): *Facing Mount Kenya*. Secker and Warburg, London, (first published 1938).

——— (1966): *My People of Kikuyu.* Oxford University Press, Nairobi.

Kohler, H. P. (2001): *Fertility and Social Interaction: An Economic Perspective.* Oxford University Press, Oxford.

Lam, D. M., and J. Miron (1996): "The Effects of Temperature on Human Fertility," *Demography*, 3, 291–305.

Manski, C., and J. Mayshar (2003): "Private Incentives and Social Interactions: Fertility Puzzles in Israel," *Journal of the European Economic Association*, 1(1), 181–211.

Manski, C. F. (2000): "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14(3), 115–36.

Manski, C. F., and J. Mayshar (2002): "Private and Social Incentives for Fertility: Israeli Puzzles," Working Paper 8984, National Bureau of Economic Research.

Manski, Charles, F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

Meck, M. (1971): "Problems and Prospects of Social Services in Kenya," Weltforum Verlag, Munchen, IFO Institut fur Wirtschaftsforschung Munchen Afrika-studien 69.

Middleton, J., and G. Kershaw (1965): *The Kikuyu and Kamba of Kenya: The Central Tribes of the North-Esatern Bantu.* International African Institute, London.

Montgomery, M. R., G. Kiros, D. Agyeman, J. B. Casterline, P. Aglobitse, and P. C. Hewett (2001): "Social Networks and Contraceptive Dynamics in Southern Ghana," Population Council Research Division Working Paper No. 153, New York.

Morgan, W. T. W., and N. M. Shaffer (1966): *Population of Kenya, Density and Distribution.* Oxford University Press, Nairobi.

Moulton, B. R. (1986): "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32, 385–397.

Mullahy, J. (1997): "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *The Review of Economics and Statistics*, 79(4), 586–593.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85.

Munshi, K., and J. Myaux (2006): "Social Norms and the Fertility Transition," *Journal of Development Economics*, 80(1), 1–38.

NAUCK, B. (1995): "Regionale Millieus Von Familien in Deutschland Nach der Politischen Vereinigung," in *Familie im Brennpunkt, Neuwid*, ed. by B. Nauck, and C. Onnen-Isemann, pp. 91–121.

OMINDE, S. H. (1968): *Land and Population Movements in Kenya.* Heinemann, London.

SRINIVAS, M. N. (1994): *The Dominant Caste and Other Essays.* Oxford University Press, Oxford, Revised and enlarged edition.

THROUP, D. W. (2001): "Kenya: Revolution, Relapse or Reform?," Africa Notes, Centre for Strategic and International Studies, November, pp. 1-9.

WATKINS, S. C., N. RUTENBERG, AND S. GREEN (1995): "Diffusion and Debate: Controversy and Reproductive Change in Nyanza Province, Kenya," Paper presented at the Annual Meetings of the Population Association in America, San Francisco, April 6-8.

WERE, G. S. (1967): *A History of the Abaluyia of Western Kenya, 1500-1900.* East African Publishing House, Nairobi.

WILLIS, R. J. (1973): "A New Approach to the Economic Study of Fertility Behaviour," *Journal of Political Economy*, 81(2 pt 2), 14–64.

WINDMEIJER, F., AND J. SANTOS-SILVA (1996): "Estimation of Count Data Models with Endogenous Regressors; an Application to Demand for Health Care," Institut fur Statistik Sonderforschungbereich 386, Paper 20.

ZHEN, X. (2008): "Semiparametric Bayesian Estimation of Mixed Count Regression Models," *Economics Letters*, 100, 435–438.

Table 1: Children Ever Born (CEB) by Region and Ethnicity, Kenya DHS Sample 1993

	Region								
	Nairobi	Central	Coast	Eastern	Nyanza	Rift Valley	Western	μ^E_{CEB}	σ^E_{CEB}
Ethnicity									
Kalenjin	1.750	8.000	9.000	0.333	4.600	3.303	2.933	3.296	3.126
Kamba	1.688	3.714	2.619	2.842	2.000	3.800	-	2.729	2.799
Kikuyu	1.682	2.498	1.735	2.091	2.000	2.983	5.333	2.491	2.438
Kisii	1.308	1.667	4.000	0.000	2.716	2.625	8.500	2.687	2.803
Luhya	1.967	2.100	2.135	3.000	3.538	3.352	3.272	3.176	3.078
Luo	1.862	3.100	1.958	2.500	3.745	3.091	3.114	3.469	3.333
Meru/Embu	0.438	2.667	1.600	2.880	1.333	0.800	-	2.744	2.705
Miji/Swa	-	0.500	3.078	1.500	-	-	1.000	3.062	2.967
μ^R_{CEB}	1.692	2.512	2.865	2.829	3.310	3.224	3.271	2.983	-
σ^R_{CEB}	1.833	2.405	2.840	2.816	3.232	3.048	3.168	-	-

μ^E_{CEB} (σ^E_{CEB}) denote, respectively, mean (variance) of CEB by ethnicity.
μ^R_{CEB} (σ^R_{CEB}) denote, respectively, mean (variance) of CEB by region.
$-$ denotes a zero cell count. See Table 5.

Table 2: Mean Education (Years) by Region and Ethnicity.

	Region							
	Nairobi	Central	Coast	Eastern	Nyanza	Rift Valley	Western	μ^E_{ED}
Ethnicity								
Kalenjin	13.5	0.0	7.0	7.3	7.0	6.5	7.0	6.5
Kamba	9.2	4.9	6.9	6.7	8.5	6.6	-	6.9
Kikuyu	9.5	7.9	9.6	8.7	7.7	7.4	6.7	8.1
Kisii	10.5	6.0	5.5	12.0	6.8	6.7	6.0	6.9
Luhya	8.7	6.8	8.4	9.6	6.0	6.0	7.1	7.0
Luo	8.7	6.7	8.2	8.8	5.9	6.9	8.1	6.3
Meru/Embu	8.9	8.2	7.2	6.5	10.0	8.6	-	6.6
Miji/Swa	-	11.0	5.1	4.0	-	-	6.0	4.2
μ^R_{ED}	9.2	7.8	5.1	6.7	6.3	6.6	7.1	6.7

Note: μ^R_{ED} (μ^E_{ED}) denotes mean years of education by region (ethnicity).
$-$ denotes zero cell count.

Table 3: Children Ever Born.

Total Children ever Born	Frequency	Percentage
0	1568	26.16
1	777	12.96
2	693	11.56
3	563	9.39
4	540	9.01
5	436	7.27
6	440	7.34
7	309	5.16
8	246	4.10
9	196	3.27
10	128	2.14
11	49	0.82
12	34	0.57
13	10	0.17
14	3	0.05
15	2	0.03

Table 4: Mean Rainfall

Mean Rainfall (year)	Region							
	Western	Eastern	Coast	Rift Valley	Central	Nairobi	Nyanza	ALL
1960	159.31	49.80	50.89	63.81	66.95	60.69	116.97	81.20
1965	155.35	55.09	70.26	54.66	57.04	74.89	137.78	86.44
1970	179.88	68.03	58.77	73.32	72.37	74.76	146.17	96.19
1975	167.08	57.20	55.34	101.70	64.09	48.86	121.12	87.91
1980	133.55	74.96	59.72	79.92	60.58	82.39	108.85	85.71
1985	188.05	119.69	63.48	90.89	67.07	60.75	131.84	103.11
1990	167.44	107.35	73.48	88.82	81.10	72.77	139.19	104.31
1995	167.25	88.19	79.04	89.22	73.98	73.57	136.41	101.09
2000	117.85	47.57	75.21	71.44	40.79	44.11	120.90	73.98
2005	125.42	58.58	63.47	75.01	61.00	55.83	102.52	77.40

Note: Mean rainfall for each province is calculated from the average rainfall of each district within that province.

Table 5: Seasonal Rainfall

Seasonality (year)	Region							
	Western	Eastern	Coast	Rift Valley	Central	Nairobi	Nyanza	ALL
1960	1.49	2.35	2.17	1.84	1.88	2.34	1.39	1.92
1965	1.50	2.71	2.21	1.93	2.13	2.21	1.69	2.06
1970	1.19	2.79	1.84	2.06	1.88	2.54	1.77	2.01
1975	1.41	2.74	2.51	1.94	2.00	2.20	1.68	2.07
1980	1.38	2.80	2.36	2.12	2.50	3.00	1.54	2.24
1985	1.17	2.58	2.22	2.01	2.13	1.98	1.56	1.95
1990	1.33	2.26	1.85	1.60	1.90	2.41	1.58	1.85
1995	1.31	2.29	2.57	1.44	1.61	1.47	1.56	1.75
2000	1.28	2.91	2.52	2.02	2.28	2.25	1.59	2.12
2005	1.38	2.72	2.10	2.15	2.67	3.02	1.82	2.26

Note: The index for the seasonality of rainfall across meteorological stations in Kenya over the course of a year (Foeken, 1994), takes the difference of the averages of rainfall in the wettest three months of the year and the driest three months and scales them by the average rainfall for the year for that particular meteorological station. the resulting calculation is an index per station and for each year in which rainfall is measured. This is reported as the index of District Rainfall Seasonality (DRS). The figures above them average this for each province.

Table 6: CEB Mean Rainfall; and Rainfall Seasonality by Region

Region	CEB (1998)	Mean Rainfall (2000)	Rainfall Seasonality
Nairobi	1.69	44.11	2.25
Central	2.51	40.79	2.28
Coast	2.86	75.21	2.52
Eastern	2.83	47.57	2.91
Nyanza	3.31	120.90	1.59
Rift Valley	3.22	71.44	2.02
Western	3.27	117.85	1.28
ALL	2.98	73.98	2.12

Table 7: GMM

	$\widehat{\beta}$	$e^{\widehat{\beta}}$	*P-value*
Individual Level			
Age	0.14284	1.15355	0.00000
Age^2	-0.00182	0.99818	0.00000
Primary Education	0.03888	1.03964	0.02900
Secondary Education	-0.04555	0.95547	0.18800
Higher Education	-0.16156	0.85082	0.04000
Listens to radio	-0.00288	0.99713	0.63300
Resident since 1993	-0.10494	0.90037	0.00000
KnowPlaceAids	0.01194	1.01201	0.25100
Household Level			
Polygyny	-0.02205	0.97819	0.31600
Usual Resident	-0.01218	0.98789	0.77100
Iron roof	-0.00794	0.99209	0.57400
Tiled roof	0.01246	1.01254	0.88000
Wood floor	-0.03655	0.96411	0.42500
Tiled floor	-0.07928	0.92378	0.43900
Cement floor	-0.00718	0.99285	0.61400
Has TV	-0.00737	0.99266	0.73400
Has radio	0.01172	1.01179	0.38400
Has telephone	0.00145	1.00145	0.96800
Has bicycle	-0.00919	0.99085	0.50500
Mortality	0.04423	1.04523	0.24100
Central	-0.01253	0.98754	0.74700
Rift Valley	-0.02475	0.97555	0.42800
Nyanza	-0.01196	0.98811	0.82600
Kalenjin	0.21871	1.24447	0.00000
Kamba	0.01675	1.01689	0.80000
Kisii	0.08097	1.08434	0.66300
Luhya	0.17899	1.19601	0.02200
Meru-embu	0.04972	1.05098	0.53300
Mijikanda Swahili	0.14131	1.15178	0.09700
Luo	0.34415	1.41078	0.00000

Table 7 GMM Parameter Estimates (continued)

	$\widehat{\beta}$	$e^{\widehat{\beta}}$	*P-value.*
Exogenous Effects: Cluster Level $\overline{E}_{e,cl}$			
Kalenjin	-0.41242	0.66204	0.42200
Kamba	-0.19543	0.82248	0.85200
Kikuyu	-0.82971	0.43618	0.12600
Kisii	-0.19361	0.82398	0.76400
Meru-embu	-0.66123	0.51621	0.48600
Mijikanda Swahili	2.04896	7.75981	0.17600
Luhya	-0.61957	0.53818	0.12500
Luo	-1.90338	0.14906	0.06800
Endogenous Effects: Cluster-level $\overline{C}_{e,cl} - C_i$			
Kalenjin	0.14475	1.15575	0.00100
Kamba	0.18885	1.20785	0.00000
Kikuyu	0.17337	1.18931	0.00200
Kisii	0.14363	1.15446	0.16700
Meru-embu	0.20078	1.22236	0.00000
Mijikanda-Swahili	0.17360	1.18958	0.00100
Luhya	0.15781	1.17094	0.01500
Luo	0.11247	1.11904	0.06500
Cluster-level Controls			
Media access (radio)	-0.12069	0.88631	0.09200
Water (nat. source)	0.11340	1.12008	0.47800
Water (piped into HH)	0.01116	1.01122	0.94400
Water (public source)	-0.02017	0.98003	0.90700
Electricity	-0.05212	0.94921	0.24400