ECONOMIC STUDIES

DEPARTMENT OF ECONOMICS
SCHOOL OF BUSINESS, ECONOMICS AND LAW
UNIVERSITY OF GOTHENBURG
196

Essays on Economic Behaviour: HIV/AIDS, Schooling, and Inequality

Annika Lindskog

UNIVERSITY OF GOTHENBURG

ISBN 978-91-85169-58-0
ISSN 1651-4289 print
ISSN 1651-4297 online

Printed in Sweden,
Geson Hylte Tryck 2011

To Henrik, Klara and Lova.

Contents

Preface

No (wo)man is an island. Many people have contributed to this thesis in various ways.

I wish to thank my supervisor Dick Durevall for the encouragement, the stimulating discussions, and the guidance. Dick also co-authored the thesis papers on HIV/AIDS and we have made two SIDA reports together. It was a pleasure working with him and I learned a lot from it.

I also want to thank Arne Bigsten, who was originally my supervisor, especially for the encouragement when I initiated the thesis work. Thank you also to Ann-Sofie Isaksson, co-author of the thesis paper on preferences for redistribution, and a good friend at the Department. I enjoyed working together. Olof Johansson-Stenman was effectively our supervisor during the work on the preferences for redistribution paper – thank you. Thanks also to Gunnar Köhlin, who was a source of help and encouragement while I was working on the children's education papers.

I am also grateful for the assistance from the administrative staff at the Department, particularly from Eva-Lena Neth-Johansson. Others at the Department who have contributed a little extra to this thesis include: Måns Söderbom, whose feedback on the final seminar improved many of the papers; Ola Olsson and Lennart Flood, who have provided useful comments; Rick Wicks, whose detailed editing significantly improved many papers; and Peter Martinsson, thanks to whom I applied to the PhD programme in the first place.

A big thanks to my fellow PhD candidates at the Department – many are PhDs by now. We have had so many exciting and interesting discussions and so much fun together. Because of you, these years were more entertaining.

I would not have become the person I am had it not been for my family and my youth friends, who are still important to me. Thank you for the friendship, the support and the belief in me – especially you, Mum and Dad.

I strongly believe that it is much easier and more enjoyable to achieve a PhD if your whole life does not centre around it. Above all I want to thank Henrik, Klara and Lova for giving true meaning and much joy to my life.

Annika Lindskog, Göteborg 24 February 2011

Summary of the thesis

The thesis consists of six self-contained papers, some being more related to each other than others.

Papers 1 to 3 address the HIV/AIDS epidemic in Malawi. In the worst affected countries in Sub-Saharan Africa, HIV rates have exceeded 10% among adults for more than two decades, generating many-fold increases in prime-age mortality (Oster, 2010; UNAIDS, 2010). In Malawi, as in other countries, the epidemic first spread in the major cities and then in rural areas. The national HIV rate was estimated to 11% in 2009 (UNAIDS, 2010).

There is an ongoing debate about the drivers of the epidemic in Sub-Saharan Africa, and the **first paper** contributes to this debate. It analyzes the relationship between economic inequality and the spread of HIV among young Malawian women. In recent years economic inequality together with gender inequality have been suggested as main socioeconomic drivers of HIV (Nattrass, 2008; Krishnan et al., 2008; Whiteside, 2008, Ch. 3; Fox, 2010), and cross-country empirical evidence supports this view (Holmqvist, 2009; Tsafack Temah, 2009; Sawers and Stillwaggon, 2010a). Although useful, cross-country regressions are likely to suffer from omitted variable biases. In particular, if absolute income matters for health and there are diminishing health returns, a relationship between health and income inequality is produced at the aggregate level when individual income is not controlled for, even if income inequality has no casual effect on health (Gravelle et al., 2002; Deaton, 2003).

We estimate multilevel logistic models of young women's individual probability of being HIV infected. Two different community levels are considered: the immediate neighbourhood and Malawi's districts. The main finding is a strong positive association between communal inequality and the risk of HIV infection. The relationship between HIV status and income, at the individual and communal levels, is less clear-cut, yet individual absolute poverty does not increase the risk of HIV infection. Further analysis shows that the HIV-inequality relationship is related to riskier sexual behaviour, gender violence and close links to urban areas, measured by return migration. However, no variable completely replaces economic inequality as a predictor of HIV infections. The HIV-inequality relationship does not seem to be

related to worse health in more unequal communities. In the debate, bad health and undernourishment have been claimed to be more important intermediating factors than sexual behaviour, since they increase the per-contact transmission rate (Stillwaggon, 2006, 2009; Sawers and Stillwaggon, 2010a). Our results do not support this view. Nor do we find the HIV-inequality relationship to be related to gender gaps in education or women's market work. Different dimensions of gender inequality thus seem to have different effects on the spread of HIV.

High HIV infection and mortality rates are likely to affect economically relevant behaviour in a variety of ways. Recently the effect of HIV/AIDS on fertility has emerged as one of the key channels through which economic growth is affected. There is a strong link between reduced fertility and economic growth in poor countries via the dependency ratio. By now there is ample evidence that the physiological effects of HIV reduce fertility by about 20% to 40% (Lewis et al., 2004). Although this effect is substantial, it is limited to infected women, and the resulting impact on country-wide fertility is marginal. The evidence on behavioural changes among all women, HIV-positive and HIV-negative alike, is inconclusive, and there are many different channels through which the risk of HIV infection and the increased adult mortality could affect fertility. The **second and third papers** in the thesis contribute to this dialogue.

The **second paper** evaluates the impact of the HIV/AIDS epidemic on the reproductive behaviour of all women in Malawi, whether HIV-negative or HIV-positive, allowing for heterogeneous responses depending on age and prior number of births. A panel of yearly observations from 1980 to the survey-year was constructed for each woman, and the woman's birth history is modelled as a discrete time process with an annual binary birth/no-birth outcome. The main explanatory variable is the district HIV rate, which is allowed to have a heterogeneous effect depending on the woman's age and number of prior births. To control for the endogeneity of the spread of the HIV epidemic, district fixed effects are used. And to verify that the results are due to a behavioural response to the HIV epidemic, rather than to a biological difference in fertility between HIV-positive and HIV-negative women, information on HIV status for a sub-sample of women is used. It is found that HIV/AIDS increases the probability of a young woman giving birth to her first child, while it decreases the

probability of giving birth for older women and for women who have already given birth. The resulting change in the distribution of fertility across age groups is likely to be more demographically and economically important than changes in the total number of children a woman gives birth to.

The **third paper** studies the effect of HIV/AIDS on actual fertility in 1999-2004 and desired fertility in 2004 among HIV-negative women and men in rural Malawi, using ordered probit models. We go beyond average effects, and analyze differences in response due to gender-specific district prime-age mortality and, as in the second paper, age-specific effects. HIV has not spread randomly, and we therefore include pre-HIV district fertility to control for factors that affected fertility in the same way before and after the HIV epidemic, i.e. time-invariant factors. This proves to be important as it changes the sign of the total fertility effect from negative to positive. Actual fertility responds positively to male mortality but negatively to female mortality, while women's desired fertility responds negatively to female mortality and men's desired fertility responds negatively to male mortality. These findings are consistent with an insurance and old-age security motive for having children among rural Malawian women. When a woman risks death before her children grow up, the value of children is low, and when the risk of a husband's death is high, the value of children is high. We also find that the positive fertility response is limited to younger women, with no discernable age-pattern in desired fertility effects. Possible reasons are early marriage to reduce the risk of HIV infection and having babies early to reduce the risk of giving birth to HIV infected babies.

All three papers on HIV/AIDS in Malawi use Demographic and Health Survey (DHS) data, and combine it with district-level data from other sources. The DHS data is rich in information and contains, for example, complete birth histories of women and HIV status for a subsample of women and men. The 2004 DHS is the first nationally representative survey of HIV prevalence in Malawi.

Papers 4 and 5 are about children's primary schooling in rural Amhara in Ethiopia. More specifically, both papers estimate the effects of older siblings' literacy on primary schooling of children in the rural Amhara region during 2000-2006, using within-household variation.

Starting with an education reform in 1994 there have been dramatic changes in primary education in Ethiopia, with massive increases in enrolment, albeit from a very low starting point. More exactly, the gross primary school enrolment rate rose from 34.0% in 1994/95 to 91.3% in 2005/2006, and the net enrolment increased from 36.0% in 1999/2000 to 77.5% in 2006/07. Furthermore, the gender gap has been narrowed; the gender parity index increased from 0.6 in 1997/98 to 0.84 in 2005/2006. As is common with such large increases in enrolment, the numbers of teachers and classrooms have not increased at pace with the number of pupils, raising concerns about reduced quality (Oumer, 2009; Ministry of Education, 2005; World Bank, 2005). In Amhara the net enrolment in 2004/2005 was 54.6% for boys and 53.1% for girls. Although these rates are both lower than the country averages, Amhara is one of the few regions where net enrolment appears to be nearly as high for girls as for boys (Ministry of Education, 2005).

The data used in the two papers comes from the Ethiopian Environmental Household Survey (EEHS), collected by the Ethiopian Development Research Institute (EDRI) in cooperation with the University of Gothenburg and, during the last round, the World Bank. Four rounds of data have been collected, in 2000, 2002, 2005 and 2007. Interviews were conducted in April/May, towards the end of the Ethiopian school year, which starts in September and ends in June. The sampled households were from 13 Kebeles, i.e., villages, in the South Wollo and East Gojjam zones of the Amhara region. The two zones were chosen to represent different agro-climatic zones in the Ethiopian highlands: There is less rainfall in South Wollo than in East Gojjam. Most households in the study areas make their living from rain-fed subsistence agriculture. Access to roads and capital markets is quite limited. Most of the information on children's education was collected in the fourth round, when respondents were asked about the schooling history of all household members age 6 to 24. Data from the fourth round has been used to create annual panels on entry into first grade and primary school grade progress, for girls and boys age 6 to 16. To obtain lagged explanatory variables, these panels were complemented with data from the three previous rounds.

The **fourth paper** investigates household-level diversification of human capital investment. Returns to formal education and investment in traditional knowledge, the

alternative in a rural area in a less-developed country, are uncertain. A possible strategy for dealing with risky or uncertain returns is diversification. Such diversification should relate to risk-aversion, and be stronger in more risk-averse households. A simple model illustrating the motivation to diversify, and how this differs with risk aversion, is developed. This is followed by an empirical analysis of the effect of older siblings' literacy on school entry probability in households with heads with different levels of risk-aversion. Rural Amhara is a place with extensive informal insurance and where parents are likely to depend on children as they get old, and is hence a place where household-level diversification could be of importance. School-entry is analyzed since it is likely to be a schooling-decision where parents' views are more important than the child's preferences and revealed abilities.

Total sibling-dependency in education was found to be positive, so any diversification was dominated by other forces. But in line with diversification across brothers, the effect of older brothers' literacy was more negative (there was no positive effect) in households with the most risk-averse heads. Possible diversification across brothers, but not across sisters, has been found also in rural Tanzania (Lilleør, 2008). However, the results in the thesis paper are statistically weak and the null hypothesis of an equal effect in all households could not be rejected.

The **fifth paper** investigates the total effect of older sisters' and brothers' literacy on girls' and boys' school entry and primary school progress in rural Amhara, a place where until recently most people have had very limited experience with formal education. Theoretically there are reasons to expect both positive and negative effects of siblings' education, making the direction of a possible effect an empirical question. After the total effects of older siblings' literacy have been estimated, an attempt is made to answer which mechanisms created the effects, focusing on time-varying credit constraints and within-household spillovers affecting actual and perceived benefits and costs of schooling (siblings could for example share books, accompany each other to school, enhance each other's learning, and affect beliefs about the benefits of schooling). The total effect turns out to be positive, and time-varying credit constraints and within-household spillovers could create positive sibling-dependency, hence the focus on these two mechanisms. To differentiate between them, literate older siblings are divided into those who were still in school and those who had left

school. With time-varying credit constraints we would expect positive effects of older siblings who had left school, but negative effects of older siblings who were still in school (due to 'competition' over scarce resources). Positive within-household spillovers would be expected both if older siblings were in school and if they had left school. To evaluate the importance of everyday interactions, literate older siblings are also divided into those who were still living in the household and those who had left.

Literacy of older sisters appears to be more beneficial than literacy of older brothers, not least since it had positive effects on school entry of both boys and girls, and since it had positive effects also when the sister had left the household. The effects of literate older siblings who were still in school and of those who had left school turned out to be similar, suggesting an important role of spillovers. The positive effects on school progress are limited to same-sex siblings who were still present in the household, suggesting an important role of everyday interactions, which could probably enhance their learning. The positive effect of sisters who had left the household suggests that they fare better than illiterate ones after leaving the household, making it possible for them to help their household of origin, but possibly also serving as a good example of the benefits of schooling, especially for girls.

With the **sixth paper** we leave both Africa and the subject health and education. It deals with determinants of preferences for redistribution in 25 countries. We attempt to explain within- and between-country variation in redistributive preferences in terms of self-interest and an input-based fairness concept, i.e. the fair distribution of income is one that rewards people who contribute with certain inputs. Dworkin (1981a, b) and later Roemer (2002) distinguish between inputs for which the individual could be considered directly responsible – 'responsible inputs' – and those that are beyond the individual's control – 'arbitrary inputs' – and argue that the fair distribution should be based only on responsible inputs.

In the empirical analysis, income is used to capture the effect of self-interest, and beliefs about causes of income are used to capture the effect of the input-based fairness concept. We use the ISSP Social Inequality III survey data set from 1999/2000. Beliefs about the causes of income differences are likely to vary across societies, and similarly, judgments on the extent to which perceived income determinants are under individual control are likely to vary across countries. Hence,

the effects of holding certain beliefs on redistributive preferences are allowed to differ across countries. The results of ordered probit estimations of redistributive support suggest that both self-interest and fairness-concerns matter. While differences in beliefs on what causes income differences seem important for explaining within-country variation, they do little for explaining between-country differences. Differences in the *effects* of holding certain beliefs, however, are important for explaining between-country variation in redistributive preferences, suggesting considerable heterogeneity across societies in what is considered as fair.

References

Deaton, A., (2003) Health, Inequality, and Economic Development. *Journal of Economic Literature* 41(1), 113-158.

Dworkin, R., 1981a. What is equality? Part 1: Equality of welfare. Philosophy and Public Affairs 10, 185-246.

Dworkin, R., 1981b. What is equality? Part 2: Equality of resources. Philosophy and Public Affairs 10, 283-345.

Fox, A.M., (2010), The Social Determinants of HIV Serostatus in Sub-Saharan Africa: An Inverse Relationship Between Poverty and HIV? *Public Health Reports*, 125(Suppl. 4), 16–24.

Gravelle, H., Wildman, J., Sutton, S., (2002) Income, Income Inequality and Health: What Can We Learn from Aggregate Data? *Social Science and Medicine* 54(4), 577-589.

Holmqvist, G. (2009) HIV and Income Inequality: If There is a Link, What Does it Tell us? Working Paper No. 54, International Policy Centre for Inclusive Growth, United Nations Development Programme.

Krishnan, S., Dunbar, M.S, Minnis, A.M., Medlin, C.A., Gerdts, C.E., Padian, N.S., (2008) Poverty, Gender Inequities, and Women's Risk of Human Immunodeficiency Virus/AIDS. *Annals of the New York Academy of Sciences* 1136, 101 – 110

Lewis, J. C., C. Ronsmans, A. Ezeh, and S. Gregson (2004), "The Population Impact of HIV on Fertility in Sub-Saharan Africa", *AIDS*, 18 (suppl 2): S35-S43.

Lilleør, H.B. (2008), "Sibling Dependence, Uncertainty and Education. Findings from Tanzania", Working paper no 2008-05, Centre for Applied Microeconomtrics (CAM), University of Copenhagen.

Ministry of Education, (2005), "Education Sector Development Programme III (ESDP-III): Program Action Plan", Addis Ababa.

Nattrass, N., (2008) Sex, Poverty and HIV. CSSR Working Paper No. 220, University of Cape Town.

Oumer, J. (2009), "The Challenges of Free Primary Education in Ethiopia", International Institute for Educational Planning (IIEP), Paris.

Oster, Emily (2010) "Estimating HIV Prevalence and Incidence in Africa from Mortality Data," *The B.E. Journal of Economic Analysis & Policy* 10(1), Article 80.

Roemer, J. E., 2002. Equality of opportunity: A progress report. Social Choice and Welfare 19, 455-471.

Sawers, L., Stillwaggon, E. (2010a), Understanding the Southern African 'Anomaly': Poverty, Endemic Disease and HIV *Development and Change* 41(2), 195-224.

Stillwaggon, E., (2006) *AIDS and the Ecology of Poverty*, Oxford University Press, Oxford.

Stillwaggon, E., (2009) Complexity, Cofactors, and the Failure of AIDS Policy in Africa. *Journal of the International AIDS Society* 12, 12-20.

Tsafack Temah, C. (2009) What Drives HIV/AIDS Epidemic in Sub-Saharan Africa? *Revue d'économie du développemen*t 23(5), 41-70.

UNAIDS, (2010) Global Report: UNAIDS Report on the global AIDS epidemic 2010. Available at http://www.unaids.org/GlobalReport/Global_report.htm.

Whiteside, A., (2008) *HIV/AIDS: A Very Short Introduction*. Oxford University Press, Oxford.

World Bank (2005), "Education in Ethiopia: strengthening the foundation for sustainable progress", World Bank, Washinton DC.

Paper I

Economic Inequality and HIV in Malawi

Dick Durevall and Annika Lindskog

Department of Economics

School of Business, Economics and Law

University of Gothenburg

Abstract

To analyze if the spread of HIV is related to economic inequality we estimate multilevel models of the individual probability of HIV infection among young Malawian women. We find a positive association between HIV infection and inequality at both the neighbourhood and district levels, but no effect of individual poverty. We also find that the HIV-inequality relationship is related to risky sex, gender violence, and return migration, though no variable completely replaces economic inequality as a predictor of HIV infections. The HIV-inequality relationship does not seem to be related to bad health, gender gaps in education or women's market work.

JEL: I12.

Key words: Africa, AIDS, gender inequality, gender violence, Malawi, poverty.

1. Introduction

Poverty is typically viewed as an important driver of the HIV epidemic, and AIDS is often called a "disease of poverty".[1] However, several studies have recently shown that poor individuals are not more likely to be HIV positive than wealthy ones, and the poorest countries among the less developed ones do not have higher infection rates than other less developed countries (Gillespie et al., 2007; Piot et al. 2007; Whiteside, 2008, p. 53). Instead, economic inequality, together with gender inequality, has been suggested as the main socioeconomic drivers of the spread of HIV (Conroy and Whiteside, 2006 Ch. 3; Nattrass, 2008; Krishnan et al., 2008; Whiteside, 2008, Ch. 3; Gillespie, 2009; Fox, 2010).

The idea that income inequality and health are related is well-established. Since the beginning of the 1990s over 200 articles have been published on the topic, and though the results vary, many find a strong association between various health indicators and income inequality across countries or regions within countries (Deaton, 2003, Subramanian and Kawachi, 2004; Wilkinson and Pickett, 2006, 2009; Babones, 2008). Yet, surprisingly few studies have analyzed income inequality and HIV/AIDS and all seem to use cross-country data (Holmqvist, 2009; Tsafack Temah, 2009; Sawers and Stillwaggon, 2010a). Although useful, cross-country regressions are likely to suffer from omitted variable biases since many potentially relevant variables cannot be included. Moreover, if absolute income matters for health and there are diminishing health returns, a relationship between health and income inequality is produced at the aggregate level even though income inequality has no casual effect on health (Gravelle et al., 2002; Deaton, 2003).

We analyze the association between economic inequality and HIV infections in Malawi; one of the countries with the highest national HIV rates in the world, 11.0% in 2009 (UNAIDS, 2010). More specifically, we consider the effect of economic inequality in the community on individual-level risks of HIV infection among Malawian women aged 15-24. The statistical analysis is carried out using multilevel logistic models of the probability of being HIV infected. We combine data from the 2004 Malawi Demographic and Health Survey (MDHS) with district-level data from the 1997/98 Integrated Household and Income Survey and 1987 Population and Housing Census. Since the size of the community might affect the results, as argued by Wilkinson and Pickett (2006), two levels of community are considered; the

[1] See for example, Whiteside (2002), Fenton (2004), Stillwaggon (2006; 2009), Wellings (2006), Dzimnenani Mbirimtengerenji (2007) and Sida (2008).

immediate neighbourhood, measured by the sampling cluster used in the 2004 MDHS, and Malawi's 27 districts.

We limit our sample to young women since they are likely to have been infected recently. This alleviates the potential problem of higher mortality among the poor, affecting studies including all prime-age adults (Sawers and Stillwaggon, 2010a). There are not enough HIV infected young men to allow estimations on them. The group of young women is also of particular interest since intergenerational transmission of HIV, which is sustaining the epidemic in the long run, mainly occurs via young women.

Our main findings are that there is a strong positive association between communal inequality and the risk of HIV infection. The relationship between income and HIV status, at the individual and communal levels, is less clear-cut. There is no evidence that poorer women are more likely to be HIV positive than others, while the results for district- and communal-level income are mixed and weak.

We also evaluate potential causes of the HIV-inequality relationship, running a series of additional regressions. The relationship appears to be due to risky sexual behaviour and gender violence, which are more common in unequal societies, but not to indicators of bad health or gender gaps in education and women's market work. To some extent, the HIV-inequality relationship can be explained by high levels of return migration from urban to rural areas, which seem to affect both inequality and HIV in communities. However, no variable completely replaces economic inequality as a predictor of HIV infections.

The paper is organized as follows. Section 2 briefly reviews earlier studies of the impact of poverty and inequality on HIV/AIDS. Section 3 describes the HIV epidemic in Malawi, and Section 4 presents our estimations strategy. Section 5 first describes the HIV data and possible sample election problems, and then the explanatory variables. Section 6 reports the empirical results, and Section 7 summarizes, discusses and concludes.

2. Inequality, Poverty and HIV/AIDS: What Do We Know?

In this section we first review the empirical evidence on HIV and economic inequality, poverty, and wealth. The focus is on Sub-Saharan Africa, where HIV mainly is transmitted through sexual contacts in the general adult population.[2] We then discuss mechanisms that potentially create links between economic inequality, poverty and HIV. There are innumerable studies of the causes of the HIV epidemic in general that are not covered here; Whiteside (2008) and UNAIDS (2008) provide general reviews.

There is strong empirical evidence that income inequality is associated with HIV prevalence at the country level. Over (1998), who analyze HIV prevalence in urban areas across developing countries, was probably the first to show this. A recent contribution is Holmqvist (2009) who, apart from carrying out his own analysis, reviews a number of studies on HIV prevalence and income distribution. The Gini coefficient of income almost always has a statistically significant coefficient. Other recent studies that obtain similar results are Nattrass (2008), Tsafack Temah (2009) and Sawers and Stillwaggon (2010a). The size of the effect varies with specification, but a change from an equal society (Gini =0.4) to an unequal society (Gini=0.6) raises prevalence by 0.5 to 1 percentage point.

Studies analyzing poverty and HIV vastly outnumber those on inequality and HIV, and the findings are not as clear-cut. Cross-country analyses give mixed results when all countries (with available data) are included. When samples are restricted to developing countries, there is usually no impact of GDP per capita or poverty on the spread of HIV (Holmqvist, 2009). In fact, relatively rich African countries have higher infection rates than poor ones.

There are also various studies using individual data that challenge the view that poor individuals have a higher risk of HIV infection (Bassolé and Tsafack, 2006; Lauchad, 2007; Mishra et al., 2007; Awusabo-Asare and Annim, 2008; Fortson, 2008; Msisha et al., 2008a). Using mainly DHS data for a number of Sub-Saharan countries, they often find that wealthy individuals are more or equally likely to be HIV positive. For example, Mishra et al. (2007)

[2] The second most important channel is mother-to-child transmission of HIV, but this is not treated in our analysis – we have data on HIV status in 2004 for women over 14 years, and people born with HIV 15 years earlier had already died by then. Some infections among adults are probably due to injections with unsterilized needles and blood transfusion with infected blood. Generally these channels are believed to be of minor importance compared to heterosexual contact, although there are divergent views (Stillwaggon, 2006; Mishra et al., 2008).

find that Malawian men in the three richest wealth quintiles are about 2.5 times more likely to be infected than those in the two poorest wealth quintiles.

A possible caveat for these findings is that wealthier people might survive longer with HIV: in cross-sectional data HIV prevalence could then be higher for richer people even if the poor have higher or equal incidence rates (Gillespie et al., 2007). Lopman et al. (2007), using Zimbabwean panel data on incidence, show empirically that wealthy HIV-positive individuals have higher survival rates than poor HIV-positive individuals, particularly among men. However, summarizing the findings of Lopman et al. and two other recent panel data studies on HIV incidence (Bärnighausen et al., 2007; Hargreaves et al. 2007), there does not appear to be a systematic pattern between getting infected and individual income.[3]

To the best of our knowledge, there are only two previous studies that analyze the role of poverty at the regional level within a country; Lauchad (2007) on Burkina Faso, and Msisha et al. (2008b) on Tanzania. They measure poverty by the headcount ratio and find it to be inversely related to HIV. Hence, several studies find that income inequality matters, while most studies on income and poverty, at individual, communal and country levels, fail to find support for the hypothesis that HIV is more common among the poor.

The association between income inequality and HIV prevalence raises questions about the mechanisms involved. In the literature on the relationship between income inequality and health in general, three main hypotheses have been suggested: the absolute income hypothesis, the relative income hypothesis, and the society-wide effects hypothesis (Leigh et al., 2009).

According to the absolute income hypothesis, it is really poverty, not income inequality, which generates the relationship. A region with high average income could have bad health when there is high income inequality simply because there are many with low incomes. Additionally, if there are diminishing health returns to income, which seems likely, then an analysis of aggregate data produces a relationship between income inequality and health even though income inequality has no casual effect on health (Gravelle et al, 2002; Deaton 2003; Jen et al., 2009).

[3] Lopman et al. (2007) find that poor men, but not women, have a higher risk of HIV incidence. Bärnighausen et al. (2007) find higher HIV incidence among individuals from the middle wealth tercile than among individuals in the poorest or richest wealth tercile, analyzing data from rural KwaZulu Natal, and Hargreaves et al. (2007) find no association between wealth and HIV incidence in data from Limpopo Province in South Africa.

The relative income hypothesis states that income inequality is an indicator of social distance between individuals, and the larger the distance the more psychosocial stress and, consequently, worse health (Wilkinson and Pickett, 2006; 2009). Accordingly, an increase in income inequality can reduce health even if everybody gets a higher income. Although the relative income hypothesis is most popular in social science fields other than economics, the idea that 'utility' depends on comparisons of own income and consumption to that of others dates far back in economics (Veblen, 1899; Duesenberry, 1949). And recently it has gained empirical support through studies in behavioural economics (Luttmer, 2005; Johansson-Stenman and Martinsson, 2006; Fliessbach et al., 2007).

The society-wide effects are related to social capital, where inequality reduces trust and increases crime and violence (Leigh et al., 2009). This mechanism is related to the relative income hypothesis, since, for instance, low social status makes people feel disrespected, which in turn can generate violence (Wilkinson and Pickett, 2006). Another possible society-wide effect is lower provision of public goods (Banerjee and Somanathan, 2007).

There is little agreement on the relative importance of the three hypotheses. The reviews by Wilkinson and Pickett (2006) and the study by Babones (2008) conclude that there is ample support for the second and third hypotheses. Deaton (2003), on the other hand, argues that there is no direct link to ill health from income inequality. The empirical findings are due to factors other than income inequality per se, poverty being one explanation. And Jen et al. (2008; 2009) obtain support for the diminishing health returns to income hypothesis. It is also possible that a third factor affects both income inequality and health. Differences in patience (discount rates) could affect investments in both education (determining income) and health. Leigh et al. (2009) go even further, arguing that the relationship between income distribution and health is fragile or non-existent. However, they base their argument only on 'robustly estimated panel specifications' which might be too demanding if a change in inequality affects health with a long lag (Deaton, 2003, Glymor, 2008). Subramanian and Kawachi (2004) take middle view, arguing that the results are inconclusive, although inequality seems to matter in unequal societies such as the U.S.

Since HIV primarily is transmitted through sexual intercourse, the potential mechanisms that relate income inequality to the spread of HIV might differ from those relevant for health in general. The main behavioral, proximate, driver of the HIV epidemics in Eastern and Southern Africa is believed to be the habit of having concurrent sexual partners and/or risky sex in general (Halperin and Epstein 2004; Whiteside, 2008, Chap. 3; Mah and Halperin, 2010). The importance of concurrent partnership is not accepted by all researchers, however.

For instance, Sawers and Stillwaggon (2010b) argue that the empirical support is weak or non-existent, and Mapingure et al. (2010) fail to find that the number of sexual partners matters when comparing samples from Tanzania and Zimbabwe. Instead, bad health and undernourishment are claimed to be more important intermediating factors, since they increase the per-contact transmission rate (Stillwaggon, 2006, 2009; Sawers and Stillwaggon, 2010a). There is, for example, strong evidence that other sexually transmitted diseases, such as genital herpes, increase the risk of HIV transmission and that malaria increases the viral load in HIV positive people (Abu-Raddad 2006; Beyrer, 2007).

The absolute income hypothesis is relevant for HIV/AIDS, since there is agreement that low income is related to poor health status in less developed countries, (Wilkinson and Pickett, 2006). There are also good reasons to expect poverty to increase the risk of HIV infection. As mentioned, bad health is one reason. Another one is that poverty is believed to make people short-sighted, and therefore more likely to take risks, since they care little about what happens to them ten years later (Oster, 2007). Women may exchange sex for goods or money to stay above the subsistence level. And men, who often have to leave their families for extended periods to work far away from home, may engage in extra marital affairs. Furthermore, poor people are more vulnerable to external shocks, such as drought, and the combined effect of poverty and shocks may increase risky behaviour substantially (Bryceson and Fonseca, 2006). The absolute income hypothesis is thus a potentially relevant explanation for the observed cross-country relationship between income inequality and HIV prevalence in Sub-Saharan Africa. With individual-level data it is possible to control for this possibility by allowing a non-linear effect of individual income.

It is also possible that a high level of poverty in a society increases infection risks for all, not only for the poor. If there is sexual networking between richer and poorer people, risky sex or undernourishment could interact with transactional sex, putting both the poor and the non-poor at greater risk of being infected. This would not be captured by individual-level income, and could be the reason why studies fail to find that poverty matters: an analysis using the level of income in the community would, however, capture the effect.[4]

The main direct link between income inequality and HIV is likely to be through transactional sex. In more unequal societies, relatively poor women may have sexual relationships because

[4] Community level income could also capture a relative income effect. Conditional on individual-level income, a higher community level income means that the individual is relatively poor, and a lower that she is relatively rich.

of aspirations to 'live a better life', not necessarily to secure the survival of themselves and their children (Fox, 2010). Even in a country as poor as Malawi, Tawfik and Watkins (2007) find that women in rural areas engage in transactional sex, not mainly to secure subsistence living, but for attractive consumer goods.[5] Moreover, in unequal societies there are likely to be more wealthy men that can afford transactional sex. If high inequality increases transactional sex, the risk of HIV will be higher for all in the sexual network.

Economic inequality could also increase the spread of HIV because of society-wide effects, notably due to lack of social cohesion (Barnett and Whiteside 2002, pp. 88-97). This could occur because it is difficult to mobilize collective action to implement effective responses to the epidemic in places with little social cohesion (Epstein, 2007, pp. 160-1). There could also be more gender violence in more unequal societies, since there is more violence in general, which tends to increase early sexual debut of women, as well as the number of rapes (Wilkinson and Pickett, 2006). Nonetheless, the concept social capital is multifaceted and can thus affect HIV prevalence through a number of mechanisms, as noted by Pronyk et al. (2008) who report that social capital is associated with protective psychosocial attributes and risk behaviour but with higher HIV prevalence in a study of poor rural households in South Africa.

Additionally, a relationship between inequality and HIV could exist because inequality is associated with more mobility, which seems to increase the spread of HIV (Oster, 2009). The most unequal societies in Sub-Saharan Africa tend to have an economic structure with large commercial farms and mines that generate geographical labour mobility. Prostitution and transactional sex relationships are common at these places, and it is well-known that infection rates are high among migrant workers, and that they might bring the disease to their home communities (Hargrove, 2008).

3. HIV/AIDS in Malawi

Malawi's first AIDS case was diagnosed in 1985, and from then on the epidemic spread rapidly, first in the major cities, and then in rural areas.[6] According to the most recent

[5] Apart from aspiring to 'live a better life', women have extra-marital affairs because of passion or to revenge on unfaithful husbands (Tawfik and Watkins, 2007).

[6] See Arrehag et al. (2006) and Conroy and Whiteside (2007) for more extensive descriptions of HIV/AIDS Malawi.

estimate, the national rate was 11% in 2009, which means Malawi registers the ninth highest HIV prevalence in the world (UNAIDS, 2010).

There are two main sources of information on HIV prevalence in Malawi, the 2004 MDHS and sentinel surveillance at antenatal clinics (ANCs). While the 2004 Malawi DHS are likely to provide good estimates of the prevalence rates in 2004 at the national level, the ANC data is the only systematic information available of how the epidemic has evolved over time. UNAIDS uses the ANC data to estimate annual HIV rates, which are reported for selected years between 1990 and 2007 in Table 1. The prevalence rate rose from about 2% in 1990 to close to 14% at the end of the 1990s. During the 2000s, there was a decline to 11%, which indicates that at least prevalence is not increasing.

The relatively constant level of prevalence rate during the last 10 years hides very different geographical developments: the rates are declining in urban areas and increasing in rural areas. Urban HIV prevalence peaked at 26% in 1995 among women attending antenatal clinics, and then started to decline slowly. It was 17% in 2004. In the rural areas the prevalence rate reached 10.8% in 2004 (NSO and OCR Macro, 2005; Republic of Malawi, 2006). There are also large differences across districts. Prevalence rates in some districts in Southern Region, with the highest rates are as high as 20%–22%, while in Northern and Central Region they are on average 8% and 7%, respectively (National Statistical Office & ORC Macro, 2005).

Table 1: HIV prevalence rates among adults (aged 15-49) in Malawi

Estimated national prevalence rates 1990-2007							
	1990	1993	1996	1999	2002	2005	2009
	2.1	8.0	13.1	13.7	13.0	12.3	11.0
Prevalence rates in 2004 by gender and area							
	Urban	Rural	South	Central	North		
Women	18.0	12.5	19.8	6.6	10.4		
Men	16.3	8.8	15.1	6.4	5.4		
Total	17.1	10.8	17.6	6.5	8.1		
Prevalence rates in 2004 by gender and age-group							
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Women	3.7	13.2	15.2	18.1	17.0	17.9	13.3
Men	0.4	3.9	9.8	20.4	18.4	16.5	9.5
Prevalence rates in 2004 among couples by the woman's age							
	15-19	20-29	30-39	40-49			
Both are positive	3.1	7.1	9.4	4.1			
The man is positive	2.4	5.5	8.2	3.5			
The woman is positive	2.7	4.1	4.7	2.9			

Sources: UNAIDS (2008) and UNAIDS (2010) provide time series information on estimated national rates. The other information is from and NSO and ORC Macro (2005).

Furthermore, there are large age and gender specific differences. Table 1 shows that HIV prevalence among women in the age group 15-19 is 9 times higher than for men, and 3.4 times higher in the age group 20-24.

In couples it is more common that only one of the two are HIV positive than that both are so, as also seen in Table 1. It is more common that the man is the only HIV-positive partner, though the difference between men and women is not large.

Although Malawi's HIV epidemic is still unfolding, it seems to have reached a relatively mature stage. As evident from Table 1, national prevalence rates have not changed much during the last 10 years, and forecasts at the regional level indicate that the infection rates will remain stable the coming years (Geubbles and Bowie, 2007). Hence, the main drivers should have had time to affect the HIV rates across Malawi, making a cross-section analysis of a fundamentally dynamic process worthwhile.

4. Empirical model

To analyze the impact of inequality on HIV, we use a multilevel logistic model.[7] It allows us to evaluate the effect of inequality at different levels on individual risk of HIV infection while accounting for other differences across communities, including unobserved ones.

With a discrete dependent variable, such as HIV status, there are no good alternative methods to both evaluate the effect of community-level regressors and control for other differences between communities. In a linear regression model we could have included community fixed effects in a individual-level regression, and then regressed the community effects on our community-level regressors. But to include community dummies in a binary model with few observations in each community would in our case lead to biased results due to the so called incidental parameters problem (Neyman and Scott, 1948). And, with the conditional fixed effects logit we would not get estimates of the community effects.

As opposed to aggregate level analysis, we can control for individual income, allowing for a non-linear effect on the probability of HIV infection. Thus we control for the effects of individual-level absolute poverty and wealth that could otherwise be confounded with inequality. Furthermore, we include community-level income to control for possible society-wide effects of community poverty or wealth.

[7] See Gelman and Hill (2007) for a lucid description of multilevel models.

We introduce community effects at two different levels, the neighbourhood, approximated by the sampling cluster, and the district. The probability of individual *i*, living in neighbourhood *j* and district *d*, being HIV-infected is

$$\begin{aligned} \Pr\left(HIV_i = 1\right) &= logit^{-1}\left(inc_i\beta_{inc\_i} + x_i^I\beta_I + \alpha_{jd[i]}^{Neigh} + \alpha_{d[i]}^{Dist}\right) \\ \alpha_{jd[i]}^{Neigh} &\sim N\left(\beta_{inc\_n}inc\_n_{jd} + \beta_{ineq\_n}ineq\_n_{jd} + x_{jd}^I\beta_N\right) \\ \alpha_{d[i]}^{Dist} &\sim N\left(\beta_{inc\_d}inc\_d_d + \beta_{ineq\_d}ineq\_d_d + x_d^I\beta_D\right). \end{aligned} \quad (1)$$

According to Eq. (1), the individual risk of being HIV infected depends on household income, inc_i, other individual-level characteristics, x_i^I, a neighbourhood effect, $\alpha_{jd[i]}^{Neigh}$, and a district effect, $\alpha_{d[i]}^{Dist}$. The neighbourhood and district effects depend on the income level and economic inequality, other community variables, and an unexplained part. The unexplained parts of the neighbourhood and district effects are assumed to be normally distributed and independent of regressors.[8]

The assumption that the unexplained parts of the community effects are normally distributed is an improvement over assuming no community-level variation in addition to that captured by regressors, but the true variation might of course have a different distribution. As a robustness check, we therefore estimate models assuming a discrete distribution with a finite number of mass-points, where the probability that a unit belongs to a certain mass-point is estimated together with its locations.

Another potential concern is that the unexplained part of the community effect is assumed not to be correlated with the regressors. If we had used only individual-level regressors this assumption would certainly be problematic: it is difficult to argue that individual poverty or wealth is not related to community characteristics that could matter for the spread of HIV. However, we assume individual-level poverty or wealth to be independent on community factors relevant for the spread of HIV *conditional* on community covariates, including the wealth of a typical household and economic inequality, a far less problematic assumption in our view. Still, as an additional check, we also estimate a model with fixed district effects, using district dummies.

[8] The likelihood functions adherent to Eq. (1) is solved by numerical approximation using adaptive quadrature. More quadrature points gives better estimates but is more computationally demanding. To ensure that we use enough quadrature points we first estimated the model using 8 points and then 15 points. If the increase in quadrature points has no substantial effect on the log-likelihood value and the estimated parameters, we have enough quadrature points. A suggested rule of thumb is that the parameter should change with less than one percent.

Our dependent variable is HIV status. We know if an individual is HIV positive, but not when he or she was infected. If HIV-infected individuals who belong to certain groups survive longer than others, this could bias our parameter estimates. Thus, we restrict our sample to young women (age 15-24) who are likely to have been infected recently to make sure that our results are not influenced by differences in mortality. There are too few HIV-infected males in this age group to estimate the models, and including older men weakens the link to the neighbourhood since many of them are mobile.[9]

Logit coefficients are not very revealing about the size of the impact of covariates. Instead we present comparisons of predicted probabilities of HIV infection when the covariates of interest are set to specific values. Predicted probabilities are computed for each woman in the sample, and include the predicted unobserved effects, i.e. the predictions are made with respect to the posterior distribution of unobserved effects.

5. Data and Variables

Our main source of data is the 2004 MDHS. This is the first nationally representative survey of HIV prevalence in Malawi, and the first to link HIV status with characteristics of the respondents and their household. There are 1,202 women aged 15-24 with available HIV status information. We also use data from the Integrated Income and Household Survey 1997/98 and the census from 1987 for measures of district-level median consumption, consumption inequality and population density, and data from the 2000 MDHS for measures of district mobility.

5.1 The HIV data and possible sample selection

In the 2004 MDHS sample, one third of the households were selected for HIV testing. The result of the test was not revealed to respondents.[10] As can be expected in any survey, particularly one that collects information about potentially sensitive issues, not all selected individuals could or wanted to participate, raising questions about the representativeness of the HIV-status sample.

[9] We also estimated models with men aged 15-29. The results for district inequality are very strong while the results for neighbourhood inequality are clearly weaker than among women age 15-24. These results are available from the authors on request.

[10] The data collection team were joined by a voluntarily testing and counselling (VCT) team that offered testing for those who were interested in knowing their HIV status.

There are two main groups with missing HIV status: respondents that were not interviewed, mainly due to absence, and respondents that were interviewed but refused to provide the blood sample for HIV-testing. Out of the 1.665 selected and interviewed women aged 15-24, HIV status data was successfully collected for 72.2%.

In the final 2004 MDHS report, the issue of potential response bias is investigated by comparing observed and predicted HIV rates for different groups of people (NSO and ORC Macro, 2005).[11] In general, observed and predicted rates differ little. The exception is Lilongwe District, where HIV status data was collected from less than 40% of the selected women, and the observed HIV rate was unreasonably low in comparison to both the predicted rate and rates observed in ANC data. Because of this we exclude Lilongwe District from our analysis. We also exclude the few observations from the small island Likoma, reducing the number of observations to 1.161 young women.

With an appropriate instrument, sample selection techniques could be used to correct for possible sample selection bias. In a study on HIV prevalence in Burkina Faso, Lachaud (2006) uses the questionable instruments urban residence and employment status, and finds no sample selection bias. Janssens et al. (2009), in a study from Windhoek, Namibia, use the more convincing instrument 'nurse who collected blood samples', and find that HIV-positive individuals are more likely to refuse the test.[12] Since we cannot think of any good instrument in our data we choose not to use sample selection techniques.

What we can do, in addition to excluding observations from Lilongwe district, is to compare differences in observables between respondents that provided the blood sample and those who refused. If people refuse the test because they know or suspect that they are HIV positive and do not trust the anonymity of the test, then refusal might be related to riskier sexual behaviour or earlier HIV testing. If refusal is related to wealth this is problematic as we intend to study the impact of wealth and its distribution on the risk of being HIV infected.[13]

---

[11] Predicted rates are constructed by first regressing HIV status on a wide range of individual and household characteristics for available HIV status observations, and then predicting HIV rates based on characteristics of all observations selected for HIV testing.

[12] The study by Janssens et al. (2009) indicates that recent HIV rates estimated from population-based surveys, which are generally substantially lower than earlier estimates based on primarily ANC data, might underestimate true HIV prevalence rates. This is also that the case if absent people, who are likely to be more mobile, have a higher risk of HIV infection.

[13] Respondents from Lilongwe and Likoma were not included in this analysis.

Young women who refused the HIV test differ in some ways from those who did not (see Table A1 in the Appendix). They seem to be less sexually active, use fewer condoms, are on average married to younger men, have less education, live in somewhat poorer districts, and are less likely to report knowing someone who had or died of AIDS. However, there is no difference in terms of wealth or communal inequality, or if previously tested for HIV. Hence, there is no evidence that they are more likely to be HIV positive than those who accepted to be tested.

5.2 Explanatory variables

We measure our community variables at two different levels: the neighbourhood, approximated by the sampling cluster (roughly a village), and the district. The major cities, Blantyre – the commercial centre, Zomba – a university town in the South, and Mzuzu – 'the capital of the North', though formally part of larger districts, are treated as separate 'districts'. Lilongwe District, which includes the capital city Lilongwe, and Likoma District are excluded from the analysis as previously explained. In total we have 340 neighbourhoods and 28 'districts'.

Individual-level income is measured by the household wealth quintile, where wealth quintiles are based on a wealth index created using information on housing characteristics and a wide range of assets. The weights attached to each item in the index are the 'coefficients' of the first principal component in a principal components analysis. Similar wealth indices have been demonstrated to be good proxies of permanent income (Filmer and Pritchett, 2001). Neighbourhood income is measured by the wealth of the typical household, the cluster median of the household wealth index, and neighbourhood inequality is measured by the household wealth index Gini coefficient.[14] At the district level, income is measured by the median level of consumption in 1997, and inequality is measured by the consumption Gini coefficient. Consumption is generally viewed as a good measure of permanent income. The variables are from the Integrated Household Survey 1997-98 published in National Economic Council (2000) and NSO (2000), respectively.[15] One advantage with using data from well before 2004 is that the simultaneity problem is reduced since there cannot be feedback effects.

[14] We also used the distance between the household wealth indices at the 90^{th} and 10^{th} percentiles as an alternative neighbourhood inequality measure. The choice of measure does not have any impact on the results.

[15] Expenditure levels have been adjusted with 4 regional consumer price indices.

In our data from Malawi, income and inequality are correlated with population density and closeness to urban areas. People in such areas are likely to be more mobile and interact with a larger number of people, which might increase the spread of HIV. In order not to confound this possible effect with wealth and inequality, we add a number of controls at both the neighbourhood and the district level.

We use GPS coordinates of the sampling clusters to create measures of distances to road, to the closest of Malawi's four main cites, and to the most important border crossing to Mozambique (in the southeast along the main transport route). When computing the distance to road, consideration is taken to level curves, i.e. the distance around rather than across mountains is used. Distance to cities and the Mozambique border crossing is computed along roads and major paths. In DHS surveys that collect blood samples for HIV testing, a random error is added to GPS coordinates, creating measurement errors.[16] This is, however, unlikely to lead to biases in our estimates. Finally, we have an indicator of urban residence at the neighbourhood level.

At the district level we use population density in 1987 and mobility of the male population. Population density is calculated using data on district area and population from the Population and Housing Census in 1987. We have not been able to separate the three cities from their surrounding districts in creating the population density figures. The 2000 MDHS data set was used to create a district-level measure of the share of the district's male population that was mobile the previous year. A man is considered mobile if he was away throughout a whole month or at five or more different occasions during the past twelve months.

Finally, in the basic models we include dummies for the respondents' level of education, none or incomplete primary (reference category), complete primary, and complete secondary or more, and age-dummies, 15-19 (reference category), and 20-24. Education is likely to be related to income but may also capture attitudes as well as knowledge and ability to process information.

The risk of HIV infection might of course be related to a wide range of other factors, among them gender inequality, ethnicity, religion and male circumcision. However, we do not want to include more variables than necessary in our main estimations. Limiting the sample to only young women reduces it to 1,161 individuals, a fairly large number but most of these, 90%,

[16] For urban communities a random error of up to 2 km in any direction is added, and for rural communities, a random error of up to 5 km is added. To one community in each survey the random error is up to 12 km.

are HIV negative. Still, as robustness check we include individual-level indicators of all the above mentioned factors. We also try to investigate what might cause an association between inequality and HIV using indicators of sexual behaviour, health, and migratory behaviour as our dependent variables. Table A1 in the appendix provides variable definitions and summary statistics.

6. Results

6.1 Main estimations of the effect of inequality on risk of HIV infection

Results from the main estimations are reported in Table 2. Specification (1), our preferred model, is based on Eq. (1). In specifications (2) and (3) we relax the assumption that the unobserved part of the community effects is normally distributed, and approximate the distribution with discrete freely estimated mass-points: specification (2) has community effects at the neighbourhood level and specification (3) at the district level.[17] We were not able to estimate the model with community effects at both the neighbourhood and district levels; it did not converge. In specification (4) we use district dummies and normally distributed neighbourhood effects.

Table 2: Main results of HIV infection among young women: Coefficients from multilevel logistic regressions

	(1) Neighbour-hood and district effects	(2) Semi-parametric neighbourhood effects	(3) Semi-parametric district effects	(4) Neighborhood effects with district dummies
Individual- level regressors				
Age 20-24	1.816***	1.793***	1.782***	1.723***
	(0.303)	(0.298)	(0.293)	(0.283)
Second poorest	-0.0434	0.00113	-0.0608	0.0147
	(0.405)	(0.422)	(0.397)	(0.413)
Middle wealth	0.445	0.593	0.491	0.684*
	(0.373)	(0.379)	(0.366)	(0.370)
Second richest	0.539	0.787**	0.605	0.783**
	(0.378)	(0.380)	(0.371)	(0.380)
Richest	0.259	0.420	0.448	0.491
	(0.470)	(0.458)	(0.445)	(0.465)
Table 2 cont				
Primary school	-0.209	-0.295	-0.124	-0.134
	(0.354)	(0.344)	(0.341)	(0.354)

[17] When estimating specification 2 and 3 we increased the number of mass-points by one until the likelihood did not increase, i.e. until the maximum Gateaux derivative was smaller than zero.

Table 2 cont.

Secondary school	0.0567	-0.122	-0.0398	0.165
	(0.440)	(0.424)	(0.434)	(0.444)
Urban	0.192	0.416	0.212	0.209
	(0.399)	(0.409)	(0.339)	(0.417)
Constant	-6.006***	-4.854***	-3.719***	-5.235***
	(1.598)	(1.623)	(1.402)	(1.131)
Neighbourhood level regressors				
Median wealth	0.240	0.371*		0.157
	(0.203)	(0.211)		(0.217)
Inequality	4.494***	3.492**		3.211**
	(1.591)	(1.529)		(1.619)
Distance to road	-0.017	-0.013		-0.020
	(0.012)	(0.012)		(0.015)
Distance to city	0.007***	0.007**		0.005
	(0.002)	(0.003)		(0.005)
Distance to border crossing	-0.002***	-0.003***		-0.005**
	(0.001)	(0.001)		(0.003)
District-level regressors				
Median consumption	-0.201*		-0.265***	
	(0.109)		(0.101)	
Inequality	6.566**		6.090**	
	(2.711)		(2.720)	
Population density	-0.00406		-0.00279	
	(0.00266)		(0.00219)	
Male mobility	1.059		-0.596	
	(1.735)		(1.715)	
Unexplained community variance				
Cluster variance	0.115			0.000
	(0.380)			(0.000)
District variance	0.000			
	(0.000)			
Semi-parametric distribution				
Location 1st mass-point		-0.144	-2.172	
prob 1		0.975	0.122	
Location 2nd mass-point		1.929	0.301	
prob 2		0.019	0.878	
Location 3rd mass-point		16.123		
prob 3		0.007		
Observations	1097	1161	1097	1141
Log likelihood	-300.1	-330.2	-308.7	-303.0

To get a sense for the magnitude of the effects, we compute predicted probabilities of HIV infection for each individual in the sample under different scenarios. First we set neighbourhood inequality equal its mean less half a standard deviation, then we set it to its mean plus half a standard deviation. Comparing the predicted probabilities in these scenarios

we get the effect of a one standard deviation increase in neighbourhood inequality around its mean. The same procedure is repeated for district inequality, neighbourhood median wealth, and district median consumption. We also compare predicted probabilities when household wealth is set to the poorest quintile, the second poorest quintile, the middle quintile, the second richest quintile, and the richest quintile. Table 3 reports the means of the predicted probabilities and Figures 1 to 5 show the cumulative distribution functions of the probabilities under the different scenarios. The predicted probabilities are based on the preferred model (specification 1).

As Table 2 reports, the effects of inequality are statistically significant at both the neighbourhood and the district levels. This result is not altered when we estimate the distribution of the unexplained part of the community effects with discrete freely estimated mass-points (specification 2 and 3). The positive effect of neighbourhood income inequality also remains when we control for unobserved district factors with district dummies (specification 4).

An increase in either neighbourhood (Figure 1) or district (Figure 2) inequality by one standard deviation around the mean creates a clear shift to the right (towards higher risk levels) in the cumulative distribution functions of the risk of HIV infection. The increases in neighbourhood and district inequality raise the mean risk of HIV infection by 2.6 and 3.2 percentage points, respectively (Table 3). Given a mean infection rate at about 10% for the women in our sample, these effects are sizeable.

The income level in the community does not have a consistent impact on the risk of HIV infection. When measuring it by median wealth in the neighbourhood, there is no noticeable change in the risk of HIV infection as wealth increase with one standard deviation around the mean (Figure 3), and the coefficient in Table 2 is generally not statistically significant. The exception is a positive effect of higher neighbourhood income, statistically significant at the ten percent level, in the estimation with semi-parametric neighbourhood effects. However, when using median district consumption, living in a poorer district is associated with an increased risk of HIV infection (Figure 4 and Tables 2-3); the mean risk increases with 2.4 percentage points as district median consumption decreases with one standard deviation around its mean.

Household wealth does not have a consistent impact on HIV infection, indicating that absolute poverty at the individual level is not related to higher prevalence rates (Table 2-3 and Figure 5). In fact, women from households in the middle and second richest wealth quintiles appear

to have the largest risk of HIV infection, followed by women in the richest household wealth quintiles, while women in the two poorest household wealth quintiles have the lowest risk. If all women belonged to the second richest household wealth quintile (with the highest risk) rather than the second poorest one (with the lowest risk), the mean risk of HIV infection would increase with as much as 4.3 percentage points (Table 3). However, the difference compared to the poorest group is only significant in some specifications (Table 2).

Turning to the other control variables, women aged 20-24 have a higher risk of HIV infection than women 15-19. More education does not appear to be related to a different risk of HIV infection when household wealth is controlled for. Urban residence is associated with a higher risk of HIV infection, but this effect is not statistically significant when neighbourhood distance measures are included. Living closer to the Mozambique border crossing along the main transport route in the southeast increases the risk of HIV infection, and, surprisingly, women who live closer to any of the four cities have a lower risk of HIV infection, but this is when we control for urban residence and other neighbourhood distance measures.[18] We do not find any statistically significant effects of population density or mobility of the district's male population.

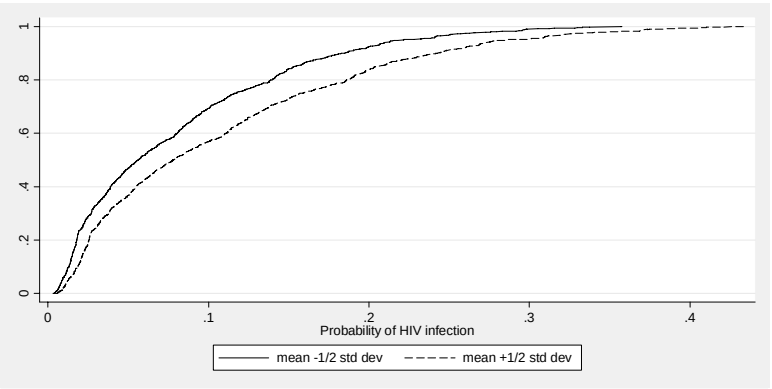
Table 3: Means of predicted probabilities of HIV infection when we change the level of an explanatory variable

	Mean
Neighbourhood inequality at its mean - 0.5 std. dev.	0.083
Neighbourhood inequality at its mean + 0.5 std. dev.	0.109
District inequality at its mean - 0.5 std. dev.	0.082
District inequality at its mean + 0.5 std. dev.	0.114
Neighbourhood median wealth at its mean - 0.5 std. dev.	0.089
Neighbourhood median wealth at its mean + 0.5 std. dev.	0.103
District median consumption at its mean - 0.5 std. dev.	0.107
District median consumption at its mean + 0.5 std. dev.	0.084
Household wealth quintile=Poorest	0.079
Household wealth quintile=Second Poorest	0.076
Household wealth quintile=Middle	0.111
Household wealth quintile=Second richest	0.119
Household wealth quintile=Richest	0.097

Note: Predicted probabilities of HIV infection, for each individual in the sample, were computed based on Specification 3 in Table 1.

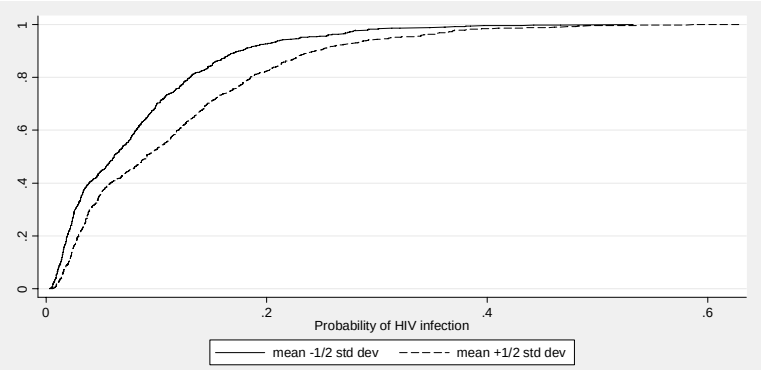
[18] This result disappears when the distance to the Mozambique border crossing is dropped.

Figure 1: The effect of neighbourhood inequality on the risk of HIV infection (cumulative distribution functions of predicted probability of HIV infection).



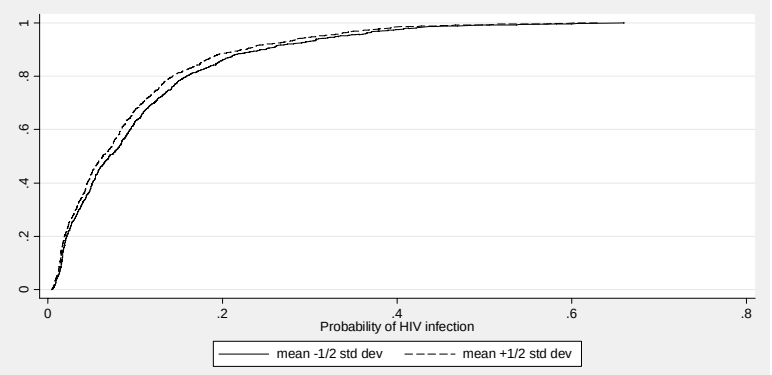
Note: Predicted probabilities of HIV infection, for each individual in the sample, were computed based on specification 3 in Table 2.

Figure 2: The effect of district inequality on the risk of HIV infection (cumulative distribution functions of predicted probability of HIV infection).



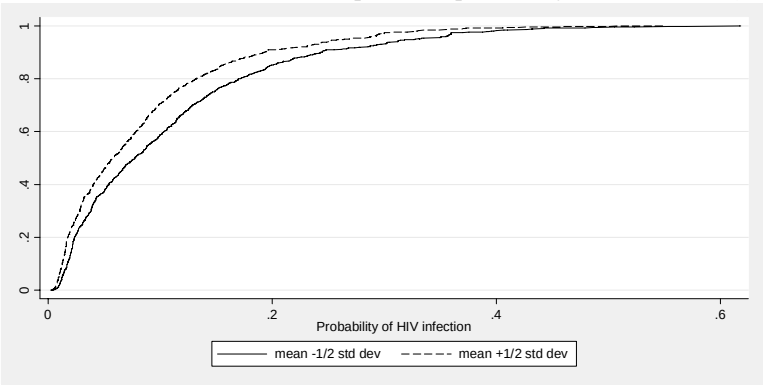
Note: See Figure 1.

Figure 3: The effect of neighbourhood median wealth on the risk of HIV infection (cumulative distribution functions of predicted probability of HIV infection).



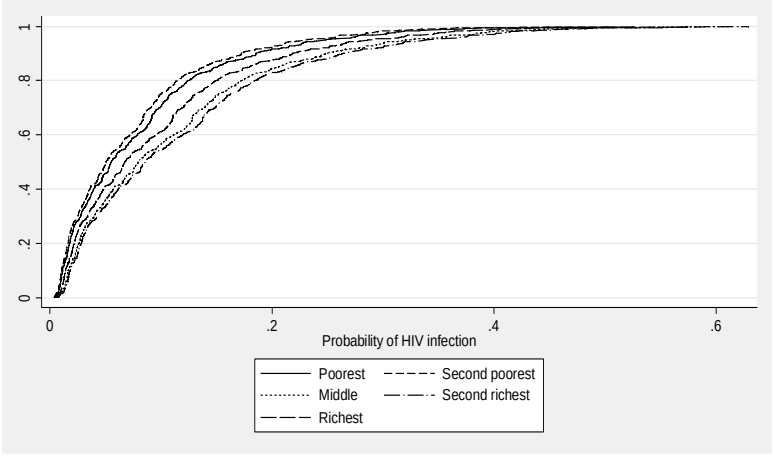
Note: See Figure 1.

Figure 4: The effect of district median consumption on the risk of HIV infection (cumulative distribution functions of predicted probability of HIV infection).



Note: See Figure 1.

Figure 5: The effect of household wealth on the risk of HIV infection (cumulative distribution functions of predicted probability of HIV infection).



Note: See Figure 1.

6.2 Why is inequality associated with an increased risk of HIV infection?

In this section we first investigate whether the association between HIV infection and inequality can be related to differences in sexual behaviour, general health, or return migration. Then we check if the results in Table 2 are robust to the inclusion of a number of other potential drivers of HIV in our model.

Table 4 report multi-level regressions with five different sexual behaviour indicators as dependent variables. Since young women's risk of HIV infection not only is affected by their own behaviour, but also by that of their sexual partners and others in a common sexual network, we also consider men's and older women's sexual behaviour when appropriate. Reporting bias is likely to be a serious issue in survey data on sexual behaviour, but we do not

see any reason why it should be systemically related to inequality or wealth. The consequence should then be a classical measurement error problem with probable attenuation bias.

Table 4: Effect of inequality and income on sexual behaviour – Multilevel regressions

	(1)	(2)	(3)	(4)	(5)
Dependent variable	Non-spouse partners	Non-spouse partners	Non-spouse partners	Never had sex	Condom use non-spouse
Method	Ordered logit	Ordered logit	Ordered logit	Logit	Logit
Sample	Young women	Women	Men	Young women	Young women
Second poorest	-0.416**	-0.681***	-0.127	-0.270*	-0.786
	(0.179)	(0.135)	(0.185)	(0.153)	(0.480)
Middle wealth	-0.392**	-0.646***	-0.0176	0.0558	0.331
	(0.177)	(0.132)	(0.183)	(0.148)	(0.399)
Second richest	-0.166	-0.697***	0.0489	0.273*	0.194
	(0.169)	(0.133)	(0.182)	(0.145)	(0.369)
Richest	-0.096	-0.641***	0.0179	0.670***	0.583
	(0.199)	(0.159)	(0.214)	(0.171)	(0.418)
Neighbourhood median wealth	0.249***	0.272***	0.163*	0.107	-0.134
	(0.092)	(0.078)	(0.091)	(0.091)	(0.155)
Neighbourhood inequality	1.379*	2.225***	1.565**	-0.245	-0.94
	(0.823)	(0.674)	(0.790)	(0.705)	(1.809)
District median consumption	-0.153***	-0.155**	-0.017	0.0238	0.018
	(0.059)	(0.067)	(0.061)	(0.066)	(0.112)
District inequality	3.321**	2.773*	4.932***	-3.309**	2.76
	(1.436)	(1.547)	(1.497)	(1.629)	(2.785)
Observations	4514	10223	2830	4513	452
Log-likelihood	-1467.4	-2447.9	-1470.9	-1669.0	-243.9
Effect of a one standard deviation increase in inequality around the mean (probability of a positive outcome or age in years)					
Neighbourhood inequality	0.004 [3.3%]	0.004 [5.6%]	0.010 [5.8%]	-0.001 [-0.4%]	-0.013 [-4.2%]
District inequality	0.014 [15.2%]	0.004 [5.7%]	0.027 [14.8%]	-0.009 [-4.1%]	0.032 [10.9%]

All specifications also include controls for age, education and urban residence at the individual level, distance to road, city and main border crossing at the neighbourhood level, and population density and mobility of the male population at the district level. They also control for unobserved neighbourhood and district effects.
Standard errors in parentheses. Percentage changes in brackets.
*p<0.10, **p<0.05, ***p<0.01

The first three specifications are multi-level ordered logistic estimations of the number of sexual partners others than the spouse during the last 12 months, for young women (age 15-24), for all women (age 15-49), and for all men (age 15-54). There are three possible categories; 0, 1, and 2 or more. Inequality, at both the neighbourhood and district levels, is associated with a larger number of non-spousal sexual partners for both women and men; the probability that a young women reports non-spousal sex rises by 3.3% if neighbourhood inequality increases by one standard deviation around its mean, and with 15.2% if district

inequality does so. For all women, the probability of non-spousal sex rises by 5.6% when neighbourhood inequality increases and by 5.7% when district inequality increases, while men's probability of reporting non-spousal sex rises with 5.8% and 14.8%, respectively.[19] We also see an effect of household poverty on women's behaviour; the very poorest seem to have had more non-spousal partners than others. However, this is not reflected in higher rates of HIV infections. Surprisingly, women, and to a some extent men, in richer neighbourhoods have more non-spousal partners than those in poor neighbourhoods, while women in poor districts have more non-spousal partners.

Specifications (4) and (5) are multilevel logistic estimations on abstinence and condom use at last non-spousal sexual encounter among young women. Abstinence means never having had sex. District inequality, but not neighbourhood inequality, is associated with a statistically significant smaller probability of abstinence, i.e. with an earlier sexual debut. The probability of abstinence decreases with 4.1% when district inequality increases with one standard deviation around its mean. And again, poverty appears to be related to riskier sexual behaviour since abstinence is less common among the poor quintiles. Women in more unequal districts seem to use condoms more frequently, but the effect is not statistically significant.[20]

Table 5 reports specifications with health indicators and return migration as the dependent variables in multilevel regressions. If inequality is associated with worse health, increased transmission rates among unhealthy populations could be one explanation for the impact of inequality on HIV. We use two indicators of general health, both closely related to undernourishment; anaemia and stunting (children's height for age). Specification (1) is a multilevel logistic estimation of anaemia among HIV-negative women. At later stages, HIV often leads to anaemia, which is why we reduce the sample to only uninfected women. Specification 2 is a multilevel linear estimation of height-for-age, measured as the Z-score of children age 0-4. Rather than child characteristics, we include characteristics of the mother (age and level of education) in these specifications. Inequality is not associated with worse health when measured by anaemia. When health is measured by stunting, inequality has a

[19] We run similar regressions with the total number of partners, including spouse, with weaker results (not reported, but available from the author on request).

[20] We also ran a regression with condom use at last sexual intercourse, whether it was with spouse or someone else. District inequality was then associated with statistically significant higher condom-use (not reported), probably because women have more non-spousal partners in the more unequal districts and condom use is rare among spouses.

negative impact but it is not statistically significant. The effect of inequality on HIV does thus not seem to be mediated through health in general.

Table 5: Effect of inequality and income on health and return migration – Multilevel regressions

	(1)	(2)	(3)	(4)
Dependent variable	Anaemia	Height for age	Return migration	Return migration
Method	Logit	Linear regression	Logit	Logit
Sample	HIV negative women	Children age 0-4	Women in rural areas	Men in rural areas
Second poorest	0.0153	-52.73	0.0734	-0.280
	(0.150)	(94.81)	(0.163)	(0.381)
Middle wealth	-0.00463	-61.33	0.137	0.129
	(0.150)	(94.27)	(0.160)	(0.350)
Second richest	-0.0513	-55.60	0.275*	0.119
	(0.154)	(99.45)	(0.157)	(0.357)
Richest	-0.316	-129.6	0.552***	1.118***
	(0.201)	(130.7)	(0.177)	(0.394)
Neighbourhood median wealth	0.0207	-9.672	0.833***	0.416
	(0.127)	(85.88)	(0.186)	(0.400)
Neighbourhood inequality	0.721	-103.3	2.086***	3.225**
	(0.730)	(479.8)	(0.711)	(1.556)
District median consumption	0.0752	-21.83	-0.0341	-0.372***
	(0.0746)	(32.38)	(0.0851)	(0.116)
District inequality	0.0203	-345.1	1.546	4.550
	(1.838)	(828.5)	(1.991)	(2.841)
Observations	2001	7802	8996	2440
Log-likelihood	-1339.4	-72681.9	-1886.9	-476.1
Effect of a one standard deviation increase in inequality around the mean (probability of a positive outcome or Z-score)				
Neighbourhood inequality	0.004 [1.1%]	-6.321 [1.1%]	0.003 [4.4%]	0.008 [14.9%]
District inequality	0.000 [0.0%]	-17.399 [2.9%]	0.002 [2.7%]	0.010 [17.0%]

All specifications also include controls for age, education and urban residence at the individual level, distance to road, city and main border crossing at the neighbourhood level (except 3), and population density and mobility of the male population at the district level (except 4). They also control for unobserved neighbourhood (except 3) and district effects (except 4). 2 and 3 use age and education of the mother rather than the children.
Standard errors in parentheses. Percentage changes in brackets
*p<0.10, **p<0.05, ***p<0.01

Specifications (3) and (4) are multilevel logistic estimations of return migration for women and men in rural areas, where return migration is measured by a dummy variable equalling 1 if the respondent migrated from an urban to a rural area during the last five years. HIV prevalence is higher among return-migrants than in the rest of the population, since many

moves to their home villages when they fall ill in AIDS.[21] People returning from the city are also often wealthy compared to others in the village, and return migration could thus cause both the spread of HIV and inequality. It is of course questionable if people already ill in AIDS contribute much to further spread of the disease. In any case, high return migration is likely to be related to stronger urban links, which can increase both the spread of HIV and inequality. We find inequality to be associated with more return migration especially among men: a one standard deviation increase in neighbourhood inequality around its mean increases the probability of being a return migrant with 4.4% among women and 14.9% among men.

Can temporary migration and links to cities explain the full effect of inequality on risk of HIV infection among the young Malawian women? To evaluate this we add community return migration as an additional control (the share of return migrants both in the neighbourhood and in the district) to specification 1 in Table 2, i.e. where we estimate risk of HIV infection. The impact of inequality only becomes somewhat weaker (Table 6, specification 1): at the district level, the inequality effect shrinks from 3.2 to 2.8 percentage points, but the coefficient is still statistically significant at the five percent level. At the neighbourhood level the inequality effect decreases from 2.6 to 2.4 percentage points, while the coefficient now is significant at the five, instead of one, percent level.

Table 6: The effect on the inequality health relationship from adding more explanatory variables.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Regression coefficients for the added variable(s) and neighbourhood and district inequality								
Neighbourhood return migration	1.791							
	(1.945)							
District return migration	-4.203							
	(5.351)							
Ethnicity dummies		Yes						
Religion dummies		Yes						
Neighbourhood male circumcision			0.622*					
			(0.352)					
Women's market work (cluster)				-0.419				
				(0.575)				
Women's market work (district)				0.610				
				(1.108)				
Secondary education gender gap (district)					-0.155			
					(0.346)			

[21] In the 2004 MDHS data, male HIV prevalence is higher among return migrants than among other men, and female HIV prevalence is higher among return migrants than among rural women or women who migrated to cities, but not higher than among urban women (Durevall and Lindskog, 2009).

Table 6 cont.

Father beat mother						0.725**		
						0.299		
Neighbourhood ethnic diversity							0.790	
							(0.563)	
Poor*neighbourhood inequality								-0.281
								(0.420)
Poor*district inequality								2.415
								(4.848)
Neighbourhood inequality	4.160**	4.718***	4.235***	4.744***	4.786***	4.963**	4.184***	4.868***
	(0.167)	(1.566)	(1.531)	(1.631)	(1.684)	(2.102)	(1.594)	(1.725)
District inequality	5.681**	5.585**	5.367**	5.898**	5.948*	4.953	6.779**	6.055**
	(2.834)	(2.791)	(2.698)	(2.886)	(3.132)	(3.586)	(2.728)	(2.944)
Effect of a one standard deviation increase in inequality around the mean (probability of HIV infection)								
Neighbourhood inequality	0.024	0.028	0.026	0.027	0.027	0.023	0.024	0.028
District inequality	0.028	0.028	0.028	0.029	0.028	0.019	0.033	0.030

All specifications also include controls for age, household wealth, education and urban residence at the individual level, median wealth, distance to road, city and main border crossing at the neighbourhood level, and median consumption, population density and mobility of the male population at the district level . They also control for unobserved neighbourhood and district effects.
Standard errors in parentheses.
*p<0.10, **p<0.05, ***p<0.01

Finally we control for various factors that have been suggested to matter for the spread of HIV. We add explanatory variables to specification 1 in Table 2, and report the results in Table 6. First, we control for religious affiliation and ethnicity (specification 2). Religions differ in terms of norms and traditions, and may matter for the spread of the epidemic. Cross-country studies regularly find that countries with many Muslims have lower HIV rates (Sawers and Stillwaggon, 2010a). Ethnicity might also affect infection rates, most obviously since some cultural traditions involve sex (Malawi Human Rights Commission, 2006). There appears to have been a decrease in sex-related cultural practices to reduce the risk of HIV infection, but studies indicate that they still exist in Malawi (Matinga and McConville, 2004; Bryceson and Fonseca, 2006; Malawi Human Rights Commission, 2006). Moreover, ethnicity, just as religion, may be related to norms and traditions that influence sexual behaviour in general.

The inequality effects are barely affected by the inclusion of religion and ethnicity. However, religious affiliation seems to matter for the risk of HIV infection. Women belonging to the Presbyterian Church have a lower probability of HIV infection than catholic women. Muslims do not, as is the case when countries are compared, have a lower risk of HIV infection. The

effects of ethnicity on risk of HIV infection are generally not of importance when religion is controlled for.[22]

In specification (3) we add the share of men that are circumcised in the neighbourhood. Male circumcision is not negatively related to HIV infection, as would be expected from the findings of cross-country studies and controlled experiments (Auvert et al. 2005; Bailey et al. 2007; Gray et al. 2007). Indeed we find a higher risk of HIV infection in areas with a higher prevalence of male circumcision. Including circumcision only reduces the inequality effects marginally. This is probably due to the type of circumcision practiced in Malawi, which differs from the one used in many other countries and controlled experiments.

Gender inequality is often considered an important driver of HIV (Dunkle and Jewkes, 2007; Gillespie, et al., 2007; Whiteside, 2007; Andersson et al., 2008), and may well be related to economic inequality. In specifications (4)-(6) we add gender inequality, differentiating between economic gender inequality, measured by women's participation in market work and the district gender gap in secondary schooling, and gender violence, which is measured by an indicator of whether the respondent's father ever beat the mother. Economic gender inequality does not appear to increase the risk of HIV infection, and it only affects the coefficient on inequality marginally. Gender violence, on the other hand, does increase the risk of HIV infection. It also weakens the inequality effects, especially at the district level, where it is reduced from 3.2 to 1.9 percentage points.

Pongou, (2009) argues that ethnic diversity increases the spread of HIV, since in societies where infidelity is not accepted, it is easier to have more partners without being detected when ethnic diversity is high. Inequality may well be related to ethnic diversity if the distribution of wealth is more equal within than across ethnic groups, and we therefore add a measure of ethnic diversity in the community in specification (7).[23] We find no statistically significant impact of ethnic diversity and its inclusion has a negligible impact on the inequality effect.

Last, in specification (8) we add interaction terms between community inequality and a dummy indicating that the woman belong to either the poorest or the second poorest wealth quintile. The purpose is to evaluate whether inequality increases the risk of HIV infection for all young women, or perhaps only for the relatively poor ones. The interaction terms are

[22] To save space, the religion and ethnicity dummy coefficients are not reported in Table 5.

[23] Ethnic diversity is measured by the probability that two randomly drawn people will belong to different ethnic groups; see Pongou (2009) for details.

statistically insignificant; inequality appears to be bad for all women in the community. Again the pure inequality effects do not change much.

7. Summary and Concluding Remarks

The aim of this study is to evaluate the impact of inequality on the spread of HIV/AIDS. We focus on a specific high HIV-prevalence country, Malawi, and analyze how inequality at both the neighbourhood and district levels affects the individual-level risk of HIV infection. The analysis is carried out by estimating multilevel logistic models for individual women, which allow us to control both for unobserved community variation and estimate the impact of community-level explanatory variables. An advantage of modelling individual data is that aggregation problems due to heterogeneity in prevalence rates within countries are avoided. Moreover, we focus on women aged 15-24. This group is of particular interest for the intergenerational transmission of HIV, but was chosen mainly to avoid possible mortality bias, i.e. that richer people can be suspected to survive longer with HIV making prevalence rates higher for richer people even if their incidence rates are the same or lower. The main source of data is the nationally representative Malawi Demographic and Health Survey (MDHS) carried out in 2004. The district-level data was collected from various sources.

Since HIV infection, wealth, and inequality all tend to be more prevalent in urban surroundings and along transport routes, we control for the proximity of the neighbourhood to roads, cities, and the main border crossing to Mozambique using GPS information, in addition to the urban residence dummy. Furthermore, we control for population density and mobility of the male population in the district, but in contrast to most existing studies, we do not include clearly endogenous measures related to sexual behaviour and HIV knowledge.

We find a strong association between inequality and the risk of HIV infection. Although a relationship between inequality and HIV prevalence rates have been established at the cross-country level, as far as we know, this is the first study that shows such a relationship using individual-level data for a particular country. When neighbourhood inequality increases with one standard deviation around its mean, the risk of HIV infection for young Malawian women increases with 2.7 percentage points, and the effect of a similar increase in district inequality is to increase the risk of HIV infection with 3.4 percentage points. These effects are substantial, since mean levels of infection are about 10%.

So what might then explain the inequality-HIV relationship? The fact that inequality matters within Malawi suggests that inequality affects the risk of HIV infection through other channels than national policies; an explanation sometimes given for why HIV rates differ across countries (Holmquist, 2009).

There is a large literature on inequality and health in general, where some argue that the aggregate relationship is not casual but due to an effect of absolute income on health, where the health returns to income are diminishing (Deaton, 2003). In all of our estimations we control for household wealth as well as individual education, allowing for non-linear relationships. Absolute poverty does not increase the risk of HIV infection for the women in our sample, and, since we control for it, poverty or diminishing health returns cannot explain the inequality HIV relationship in our study.

Differences in heterosexual transmission of HIV depend on either differences in sexual behaviour or differences in per-contact transmission rates. If inequality is related to generally worse health, it increases per-contact transmission rates; less healthy HIV positive people may be more infectious due to higher viral loads and less healthy HIV-negative people may be more vulnerable to infection. We therefore consider the possibility of an inequality impact on two measures of the general health situation; anaemia among HIV negative women (aged 15-49) and height for age of children under age of 5. We do not find that inequality affects health in general; there is no (measurable) impact on anaemia among HIV negative women or children's height for age.

However, we do find that inequality affects sexual behaviour. It increases the probability that young women, women in general, and men report non-spousal sex during the last year. District, but not neighbourhood, inequality decreases the probability of abstinence among young women, i.e. it is related to an earlier sexual debut. However, condom use is more frequent in more unequal places, possibly because of more non-spousal sex.

It seems reasonable to assume that economic inequality is related to more transactional sex. Women not necessarily engage in transactional sex to ensure survival for themselves and their children, but since they desire the material life they see others have. And in unequal places there are relatively wealthy men that can afford transactional sex. Transactional sex is often related to concurrent partners, and the increased level of transactional sex does not only increase the risk of HIV infection for the men and women in the 'extra' transactional sex relationships, but for all in the sexual network. We have no transactional sex specific information in the data to fully confirm this hypothesis, but believe that more non-spousal sex

partners, earlier sexual debut for young women, and young women marrying older men, all are consistent with it.

Violence and lack of social cohesion are sometimes proposed to explain the relationship between inequality and health in the literature. Less social cohesion could hinder an effective response to the HIV epidemic, and gender violence is often seen as a driver of HIV. We cannot think of any good way to analyse social cohesion with our data, but controlling for gender violence, measured by the share of men and women in the community who report that their father ever beat their mother, somewhat weakens the inequality impact. Moreover, gender violence has a clearly significant effect on the risk of HIV-infection.

We also find that migratory patterns could explain part of the HIV-inequality relationship. Inequality is related to an increased presence of return migrants in rural areas, and probably also to more out-migration to urban areas, temporary and permanent, increasing contacts with the cities. It is difficult to know exactly how causation runs in the migration-inequality-HIV relationship. Inequality could increase migration, but migration could also increase inequality and HIV as migrants bring both wealth and the virus from the city back to the village. When we control for the share of return migrants in our estimations of young women's HIV infection, the impact of neighbourhood inequality is only reduced by 10-15%.

While the risk of HIV infection is not higher for women from poorer households, we do find that lower median consumption at the district level is associated with higher risks of HIV infection, and the effect is statistically significant in some specifications. The results are not that strong, and not consistent with the effect of household income.

To succeed in the long term, HIV prevention efforts need to address the underlying drivers of HIV risk and vulnerability in different settings (Geeta et al., 2008). We have focussed on one potential factor that drives the epidemic, income inequality. When our findings are combined with those of other studies, there seems to be substantial evidence that income inequality matters for the spread of HIV. Our study also shows that inequality affects the risk of HIV infection through increased sexual risk behaviour, possibly as transactional sex is more common in more unequal places, and that it is closely related to gender violence. Our results thus support measures to address economic inequality, as well as inequalities in gender relations and sexual interaction. They also suggest that part of the increased risk of HIV infection in more unequal places is due to migratory connections to cities, whether it is migration that causes inequality or inequality that causes migration. Migration to cities is likely to be favourable for economic development, and should therefore be welcomed, but

knowledge about the risks it brings could be helpful for targeting of HIV/AIDS prevention measures.

References

Abu-Raddad, L.J., Patnaik, P., Kunlin, J.G., (2006) Dual Infection with HIV and Malaria fuels the Spread of both Diseases in Sub-Saharan Africa. *Science* 314,1603-1606.

Andersson, N., Cockcroft, A., Shea, B., (2008) Gender-Based Violence and HIV: Relevance for HIV Prevention in Hyperendemic Countries in Southern Africa. *AIDS* 22(Suppl.4), S73-S86.

Arrehag, L., de Vylder, S., Durevall, D., Sjöblom, M., (2008) The Impact of HIV/AIDS on Livelihoods, Poverty and the Economy of Malawi. Sida Studies No. 18.

Auvert, B., Taljaard, D., Lagarde, E., Sobngwi-Tambekou, J., Sitta, R., & Puren, A. (2005). Randomized Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk: The ANRS 1265 Trial. PLOS Medicine, 2(11), 1112-1122.

Babones, S.J., (2008) Income Inequality and Population Health: Correlation and Causality. *Social Science & Medicine* 66(7), 1614-1626.

Bailey, R.C., Moses, S., Parker, C.B., Agot, K., Maclean, I., Krieger, J.N., et al. (2007). Male Circumcision for HIV Prevention in Young Men in Kisumu, Kenya: a Randomized Controlled Trial. *The Lancet*, 369(9562), 643–656.

Banerjee, A., Somanathan R., (2007) The Political Economy of Public Goods: Some Evidence from India, *Journal of Development Economics*, 82(2), 287-314.

Barnett, T., Whiteside, A., (2002) *AIDS in the Twenty-First Century: Disease and Globalization*. Palgrave Macmillan, New Work.

Bassolé, L., Tsafack, C., (2006) Income Inequality and the HIV/AIDS Epidemic. Presented at the IAEN (International AIDS Economic Network) meeting prior to the International AIDS 2006 Conference, held in Toronto (Canada), on 11-12 August 2006.

Beyrer, C., (2007) HIV Epidemiology Update and Transmission Factors: Risks and Risk Contexts: 16th International AIDS Conference Epidemiology Plenary. *Clinical Infectious Diseases* 44, 981-987.

Bryceson, D.F, Fonseca, J., (2006) Risking Death for Survival: Peasant Responses to Hunger and HIV/AIDS in Malawi. *World Development* 34(8), 1654-1666.

Bärnighausen, T., Hosegood, V., Timaeus, I.M., Newell, M.-L., (2007) The Socioeconomic Determinants of HIV Incidence: Evidence from a Longitudinal, Population-Based Study in Rural South Africa. *AIDS* 21(Suppl. 7), S29-S38.

Conroy, A.C., Blackie M.J., Whiteside, A., Malewezi, J.C., Sachs, J.D. (2007) *Poverty, AIDS and Hunger: Breaking the Poverty Trap in Malawi*. Palgrave Macmillan, New York.

Conroy, A., Whiteside A., (2007) The AIDS Crisis. Chap. 3 in Conroy, A., et al., Palgrave Macmillan, New York.

Deaton, A., (2003) Health, Inequality, and Economic Development. *Journal of Economic Literature* 41(1), 113-158.

Dzimnenani Mbirimtengerenji, N., (2007) Is the HIV/AIDS Epidemic an Outcome of Poverty in Sub-Saharan Africa? *Croatian Medical Journal* 48(5), 605-17.

Duesenberry, J.S., (1949) Income, Savings, and the Theory of Consumer Behaviour. Cambridge MA:, Harvard University Press.

Dunkle, K.L., Jewkes, R., (2007) Effective HIV Prevention Requires Gender-Transformative Work with Men. *Sexually Transmitted Infections* 83, 173-174.

Epstein, H., (2007) *The Invisible Cure: Africa, the West, and the Fight Against AIDS*. Penguin Books Ltd, London.

Fenton L., (2004) Preventing HIV/AIDS through Poverty Reduction: the Only Sustainable Solution? *The Lancet* 364(9440), 1186-1187.

Filmer, D., Pritchett, L.H., (2001) Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrolments in States of India. *Demography* 38(1), 115-132.

Fliessbach, K.,Weber, B..Trautner, P., Dohmen, T., Sunde, U., Elger, C.E., Falk, A., (2007), Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum. *Science* 318 (5854), 1305-1308.

Fortson, J.G., (2008), The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS. Demography 45(2), 303–322.

Fox, A.M., (2010), The Social Determinants of HIV Serostatus in Sub-Saharan Africa: An Inverse Relationship Between Poverty and HIV?. *Public Health Reports*, 125(Suppl. 4), 16–24.

Gelman, A., Hill, J., (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge, Cambridge University Press.

Geubbels, E., Bowie, C., (2006) Epidemiology of HIV/AIDS in Adults in Malawi. *Malawi Medical Journal* 18(3), 99-121.

Gillespie, S., (2009), Poverty, Food Insecurity, HIV Vulnerability and the Impacts of AIDS in sub-Saharan Africa. *IDS Bulletin* 39(5), 10-18.

Gillespie, S., Kadiyala, S., Greener, R., (2007), Is Poverty or Wealth Driving HIV transmission? *AIDS* 21(Suppl. 7), S5-S16.

Glymour, M.M., (2008), Sensitive periods and first difference models: Integrating etiologic thinking into econometric techniques: A commentary on Clarkwest's 'Neo-materialist theory and the temporal relationship between income inequality and longevity change'. *Social Science & Medicine*, 66(9) 1895-1902.

Goldstein, H., Browne, W., Rasbash J., (2002) Partitioning Variation in Multilevel Models. *Understanding Statistics* 1(4), 223-231.

Gravelle, H., Wildman, J., Sutton, S., (2002) Income, Income Inequality and Health: What Can We Learn from Aggregate Data? *Social Science and Medicine* 54(4), 577-589.

Gray, R.H., Kigozi, G., Serwadda, D., Makumbi, F., Watya, S., Nalugoda, F., (2007). Male Circumcision for HIV Prevention in Men in Rakai, Uganda: a Randomised Trial. *The Lancet*, 369(9562), 657–666.

Halperin, D.T., Epstein H., (2004) Concurrent Sexual Partnerships Help to Explain Africa's high HIV Prevalence: Implications for Prevention. *The Lancet* 364(9428), 4-6.

Hargreaves, J.R., Bonell, C., Morison, L.A., Kim, J.C., Phetla, G., Porter, J., Watts, C., Pronyk, P.M., (2007) Explaining Continued high HIV Prevalence in South Africa: Socioeconomic Factors, HIV Incidence and Sexual Behaviour Change Among a Rural Cohort, 2001-2004. *AIDS* 21(Suppl. 7), S39-S48.

Hargrove, J., (2008) Migration, Mines and Mores: the HIV Epidemic in Southern Africa. *South African Journal of Science* 104(1 & 2), 53-61.

Holmqvist, G. (2009) HIV and Income Inequality: If There is a Link, What Does it Tell us? Working Paper No. 54, International Policy Centre for Inclusive Growth, United Nations Development Programme.

Janssens, W., van der Gaag, J., Rinke-de Wit, T., (2009) Non-Response Bias in the Estimation of HIV Prevalence. Paper presented at the CSAE Economic Development for Africa Conference 2009.

Jen, M.H., Jones, K., Johnston, R., (2008) Compositional and Contextual Approaches to the Study of Health Behaviour and Outcomes: Using Multi-level Modelling to Evaluate Wilkinson's Income Inequality Hypothesis. *Health & Place* 15(1), 198-203.

Jen, M.H., Jones, K., Johnston, R., (2009) Global Variations in Health: Evaluating Wilkinson's Income Inequality Hypothesis Using the World Values Survey. *Social Science & Medicine* 68(4), 643-653.

Johansson-Stenman, O., Martinsson, P, (2006), Honestly, why are you driving a BMW? *Journal of Economic Behavior and Organization*, 60(2), 129-146.

Krishnan, S., Dunbar, M.S, Minnis, A.M., Medlin, C.A., Gerdts, C.E., Padian, N.S., (2008) Poverty, Gender Inequities, and Women's Risk of Human Immunodeficiency Virus/AIDS. *Annals of the New York Academy of Sciences* 1136, 101 – 110

Lachaud, J.-P., (2007) HIV Prevalence and Poverty in Africa: Micro- and Macro-Econometric Evidences Applied to Burkina Faso. *Journal of Health Economics* 26(3), 483-504.

Leigh, A., Jencks, C., Smeeding, T.M., (2009) Health and Economic Inequality. In Salverda, W., Nolan, B., Smeeding, T.M., (Eds.) *The Oxford Handbook of Economic Inequality*, Oxfor University Press, Oxford.

Lopman, B., Lewis, J., Nyamukapa, C., Mushati, P., Chandiwana, S., Gregson, S., (2007) HIV Incidence and Poverty in Manicaland, Zimbabwe: Is HIV Becoming a Disease of the Poor?" *AIDS* 21(suppl. 7), 57-66.

Luttmer, E., (2005) Neighbors as negatives: relative earnings and well-Being. *Quarterly Journal of Economics*, 120, 963-1002.

Mah, T.L., Halperin, D.T., (2010) The Evidence for the Role of Concurrent Partnerships in Africa's HIV Epidemics: A Response to Lurie and Rosenthal. *AIDS and Behavior*, Forthcoming.

Malawi Human Rights Commission (2006) Cultural Practices and Their Impact on the Enjoyment of Human Rights, Particularly the Rights of Women and Children in Malawi. Lilongwe, Malawi. http://www.malawihumanrightscommission.org/

Mapingure, Munyaradzi P. , Sia Msuya, Nyaradzai E Kurewa , Marshal W Munjoma, Noel Sam, Mike Z Chirenje, Simbarashe Rusakaniko, Letten F Saugstad, Sake J de Vla and Babill Stray-Pedersen (2010), Sexual behaviour does not reflect HIV-1 prevalence differences: a comparison study of Zimbabwe and Tanzania. *Journal of the International AIDS Society*, 2010, (13)45, doi:10.1186/1758-2652-13-45.

Matinga, P., McConville, F., (2004) Review of Cultural Beliefs and Practices Influencing Sexual and Reproductive Health, and Health-Seeking behaviour in Malawi. Paper presented at the Interantional Conference on AIDS July 11-16, 2004 in Bangkok, Thailand.

Mishra, V., Bignami-Van Assche, S., Greener, R., Vaessen, M., Hony, R., Ghys, P, Ties Boerma, J., Van Assche, A., Khan, S., Rutstein S., (2007) HIV Infection Does not Disproportionately Affect the Poorer in Sub-Saharan Africa. *AIDS* 21(Suppl. 7), s17-s28.

Mishra, V., Kottiri, B., Liu, L., Rathavuth, H., Khan, S., Opio, A., (2008) The Association between Medical Injections and Prevalent HIV Infection: Evidence from a National Sero-Survey in Uganda. DHS Working Paper No. 42, Macro International Inc, Calverton, USA.

Msisha, W.M, Kapiga, S.H., Earls, F.J.,. Subramanian, S.V., (2008a) Socioeconomic Status and HIV Seroprevalence in Tanzania: A Counterintuitive Relationship. *International Journal of Epidemiology* 37, 1297–303.

Msisha, W.M, Kapiga, S.H., Earls, F.J.,. Subramanian, S.V., (2008b) Place Matters: Multilevel Investigation of HIV Distribution in Tanzania. *AIDS*, 22(6), 741-748.

National Economic Council (2000) Profile of Poverty in Malawi, 1998: Poverty Analysis of the Malawi Integrated Household Survey, 1997-98. Lilongwe.

Nattrass, N., (2008) Sex, Poverty and HIV. CSSR Working Paper No. 220, University of Cape Town.

Neyman, X., Scott, X., (1948) Consistent Estimation from Partially Consistent Observations. *Econometrica* 16, 1–32.

NSO (National Statistical Office of Malawi) and OCR Macro (2005) Malawi Demographic and Health Survey 2004 – Final Report. Zomba and Calverton, Maryland.

NSO (2000) Malawi Integrated Household Survey 1997/98 - Statistical Abstract. Lilongwe.

Oster E., (2007) HIV and Sexual Behaviour Change: Why not Africa? NBER Working Paper Number No. 13049.

Oster, E., (2009) Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa. Mimeo, University of Chicago.

Over, M., (1998) The Effects of Societal Variables on Urban Rates of HIV Infection in Developing Countries: An Exploratory Analysis. Chap. 2 in Ainsworth, M., Fransen, L., Over, M., (Eds.) *Confronting AIDS: Evidence from the Developing World*, European Commission, Brussels.

Piot, P., Greener, R., Russell, S., (2007) Squaring the Circle: AIDS, Poverty, and Human Development. *PLoS Medicine* 4(10), 1571-75.

Pongou, R., (2009) Anonymity and Infidelity: Ethnic Identity, Strategic Cross-Ethnic Sexual Network Formation, and HIV/AIDS in Africa. Mimeo, Brown University.

Pronyk, Paul M., Trudy Harpham, Linda A. Morison, James R. Hargreaves, Julia C. Kim, Godfrey Phetla, Charlotte H. Watts, John D. Porter, (2008), Is social capital associated with HIV risk in rural South Africa? *Social Science & Medicine*, 66(9), 1999-2010.

Republic of Malawi, (2006) Report of the Malawi Triangulation Project: Synthesis of Data on Trends in the National and Local HIV Epidemics and the Reach and Intensity of Prevention Efforts. National Aids Commission, Lilongwe.

Sawers, L., Stillwaggon, E., Hertz T., (2008) Cofactor Infections and HIV Epidemics in Developing Countries: Implications for Treatment. *AIDS Care* 20(4), 488-494.

Sawers, L., Stillwaggon, E. (2010a), Understanding the Southern African 'Anomaly': Poverty, Endemic Disease and HIV *Development and Change* 41(2), 195-224.

Sawers, L., Stillwaggon, E. (2010b), Concurrent sexual partnerships do not explain the HIV epidemics in Africa: a systematic review of evidence. *Journal of International Aids Society*, 13(34) 2-23 34doi:10.1186/1758-2652-13-34.

Sida (Swedish Development Authority) (2008) Den svenska strategin: HIV/AIDS orsaker och konsekvenser. http://www.sida.se/sida/jsp/sida.jsp?d=425&a=1404 (accessed 13 December, 2008).

Stillwaggon, E., (2006) *AIDS and the Ecology of Poverty*, Oxford University Press, Oxford.

Stillwaggon, E., (2009) Complexity, Cofactors, and the Failure of AIDS Policy in Africa. *Journal of the International AIDS Society* 12, 12-20.

Subramanian, S.V., Kawachi, I., (2004) Income Inequality and Health: What Have We Learned So Far? *Epidemiogical Review* 26, 78-91.

Tawfik, L., Watkins, S.C., (2007) Sex in Geneva, Sex in Lilongwe, and Sex in Balaka. *Social Science and Medicine* 64, 1090-1101.

Tsafack Temah, C. (2009) What Drives HIV/AIDS Epidemic in Sub-Saharan Africa? *Revue d'économie du développemen*t 23(5), 41-70.

UNAIDS, (2008) 2008 Report on the Global AIDS Epidemic. Available at http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/.

UNAIDS, (2010) Global Report: UNAIDS Report on the global AIDS epidemic 2010. Available at http://www.unaids.org/GlobalReport/Global_report.htm.

Veblen, T., (1899) The Theory of the Leisure Class. 2008 republication at www.forgottenbooks.org

Wellings, K., Collumbien, M., Slaymaker, E., Singh, S., Hodges, Z., Patel, D. Bajos N., (2006) Sexual Behaviour in Context: A Global Perspective. *The Lancet* 368, 1706-28.

Whiteside, A., (2002) Poverty and HIV/AIDS in Africa. *Third World Quarterly* 23(2), 313–332.

Whiteside, A., (2008) *HIV/AIDS: A Very Short Introduction*. Oxford University Press, Oxford.

Wilkinson, R.G., Pickett, K.E, (2006) Income Inequality and Population Health: A Review and Explanation of the Evidence. *Social Science & Medicine* 62(7), 1768-1784.

Wilkinson, R.G., Pickett, K.E, (2009) Income Inequality and Social Dysfunction. *Annual Review of Sociology* 35, 493-511.

Table A1: Description of variables and summary statistics

Variable	Description	Obs.	Mean	Std. Dev.
Young women with HIV-status information				
HIV-status	1 if HIV positive; 0 if HIV negative	1161	0.100	0.300
Age 15-19	1 if age 15-19; 0 else (reference group)	1161	0.439	0.497
Age 20-24	1 if age 20-24; 0 else	1161	0.561	0.497
Poorest	1 if poorest household wealth quintile; 0 else (reference group)	1161	0.180	0.384
Second poorest	Second poorest household wealth quintile; 0 else	1161	0.201	0.401
Middle wealth	Middle household wealth quintile; 0 else	1161	0.211	0.408
Second richest	Second richest household wealth quintile; 0 else	1161	0.215	0.411
Richest	Richest household wealth quintile; 0 else	1161	0.193	0.395
No education	1 if less than complete primary education; 0 else (reference group)	1161	0.094	0.292
Primary	1 if complete primary but not complete secondary education; 0 else	1161	0.703	0.457
Secondary	1 if complete secondary education; 0 else	1161	0.203	0.403
Urban	1 if urban residence; 0 if rural residence	1161	0.153	0.360
Chewa	1 if ethnicity is chewa; 0 else (reference group)	1161	0.294	0.456
Tumbuka	1 if ethnicity is tumbuka; 0 else	1161	0.126	0.332
Lomwe	1 if ethnicity is lomwe; 0 else	1161	0.171	0.376
Tonga	1 if ethnicity is tonga; 0 else	1161	0.022	0.148
Yao	1 if ethnicity is yao; 0 else	1161	0.163	0.369
Sena	1 if ethnicity is sena; 0 else	1161	0.034	0.182
Nkonde	1 if ethnicity is nkonde; 0 else	1161	0.008	0.088
Ngoni	1 if ethnicity is ngoni; 0 else	1161	0.102	0.303
Other ethnicity	1 if ethnicity is other than above; 0 else	1161	0.080	0.272
Catholic	1 if Catholic; 0 else (reference group)	1161	0.245	0.430
Ccap	1 if Central African Presbyterian Church; 0 else	1161	0.178	0.383
Anglican	1 if Anglican Church; 0 else	1161	0.022	0.145
Baptist	1 if Baptist/Seventh day Adventist; 0 else	1161	0.056	0.230
Other Christian	1 if other Christian church than above; 0 else	1161	0.351	0.478
Muslim	1 if muslim; 0 else	1161	0.145	0.352
No religion	1 if no religion/atheist; 0 else	1161	0.003	0.059
Father beat mother	1 if respondent report that her father ever beat her mother; 0 else	846	0.297	0.457
Young women				
Never had sex	1 if the respondent never had sex; 0 else	4777	0.223	0.417
Condom with non-spouse	1 if the respondent used a condom last time she had sex with a non-spousal partner; 0 else	476	0.309	0.462
Non-spousal partners	Number of non-spousal sexual partners the last 12 months	4777	0.106	0.328
Women				
Non-spousal partners	Number of non-spousal sexual partners the last 12 months	10776	0.073	0.333
HIV-negative women				
Anaemia	1 if respondent has anaemia; 0 else	2106	0.415	0.493
Women in rural areas				
Return migration	1 if respondent migrated from an urban to a rural area during the last five years	9370	0.063	0.242

Table A1 cont.

Men				
Non-spousal partners	Number of non-spousal sexual partners the last 12 months	2953	0.243	0.698
Men in rural areas				
Return migration	1 if respondent migrated from an urban to a rural area during the last five years	2548	0.058	0.234
Young children (age 0-4)				
Height-for-age	Height-for-age Z-score multiplied by 100, where the Z-score is the child's height less the mean height for a child of that age in a reference population divided by the standard deviation of the mean height for a child that age in the reference population.	7802	-182.34	162.35
Neighbourhoods. i.e. sampling clusters				
Neighbourhood median wealth	Cluster median of the household wealth index	484	0.875	0.738
Neighbourhood inequality	Household wealth index Gini coefficient	484	0.294	0.079
Distance to road	Distance in km to road	484	10.3	13.0
Distance to city	Distance in km to a city. the closest of Lilongwe. Blantyre. Zomba or Mzuzu	484	94.6	61.9
Distance to border crossing	Distance in km to the Mozambique border in the southeast along the main transport route	484	329.7	245.4
Neighbourhood return migration	Share in cluster that migrated from an urban to a rural area the last five years (men and women are weighted with 0.5 each)	484	0.054	0.067
Neighbourhood male circumcision	Share of the men in the cluster that are circumcised	480	0.287	0.342
Neighbour. female market work	Share of women in the cluster that participate in market work. i.e. that works for a money income	478	0.306	0.266
Neighbourhood ethnic diversity	Ethnic diversity in the cluster. measured by the probability that two randomly chosen individuals have different ethnicities	485	0.409	0.250
Districts				
District inequality	1997 Gini coefficient of household's per capita consumption in the district	26	0.412	0.064
District median consumption	Median of 1997 household's per capita consumption in the district	27	8.437	1.421
District male mobility	Share of the district's male population in 2000 that spent either at least one month away from home. or were away at least at 5 different occasions the last 12 months	28	0.330	0.084
Population density	District population density in 1987	28	112.0	83.6
District return migration	Share in district that migrated from an urban to a rural area the last five years (men and women are weighted with 0.5 each)	28	0.052	0.030
District female market work	Share of women in the district that participate in market work. i.e. that works for a money income	29	0.381	0.195
Secondary education gap	Share of women with secondary education/ share of men with secondary education	26	0.400	0.499

Table A2: Comparison of observables between young women who provided blood and young women who refused

		Refused		HIV-tested		Equal means test	
	Obs.	Mean	Std.	Mean	Std.	F	p
Age	1554	19.631	0.142	19.786	0.081	0.91	0.342
Urban residence	1524	0.084	0.014	0.077	0.008	0.16	0.692
Less than complete primary education	1554	0.122	0.017	0.094	0.009	2.30	0.130
Complete primary education	1554	0.720	0.023	0.703	0.013	0.43	0.513
Complete secondary education	1554	0.153	0.018	0.203	0.012	5.45	0.020
Poorest household wealth quintile	1554	0.170	0.019	0.180	0.011	0.19	0.666
Second poorest household wealth quintile	1554	0.216	0.021	0.201	0.012	0.43	0.514
Middle household wealth quintile	1554	0.226	0.021	0.211	0.012	0.40	0.525
Second richest household wealth quintile	1554	0.221	0.021	0.215	0.012	0.06	0.803
Richest Household wealth quintile	1554	0.165	0.019	0.193	0.012	1.56	0.212
District inequality	1466	0.414	0.014	0.410	0.012	0.58	0.453
District median consumption	1490	8.226	0.272	8.541	0.289	6.94	0.014
Neighbourhood inequality	1554	0.294	0.002	0.296	0.001	0.86	0.353
Neighbourhood median wealth	1554	0.904	0.060	0.918	0.055	0.06	0.801
Catholic	1554	0.160	0.019	0.245	0.013	14.14	0.000
Central African Presbyterian Church	1554	0.188	0.020	0.178	0.011	0.19	0.660
Anglican	1554	0.028	0.008	0.022	0.004	0.48	0.490
Seventh day adventists	1554	0.061	0.012	0.056	0.007	0.13	0.714
Other Christian	1554	0.374	0.024	0.351	0.014	0.64	0.422
Muslim	1554	0.181	0.019	0.145	0.010	2.67	0.103
No religion	1554	0.008	0.004	0.003	0.002	0.79	0.375
Southern region	1524	0.546	0.122	0.499	0.114	0.82	0.373
Age difference to husband	654	5.199	0.317	5.810	0.182	2.79	0.095
Never had sex	1554	0.249	0.022	0.196	0.012	4.58	0.033
Any sex last four weeks	1554	0.410	0.025	0.469	0.015	4.29	0.038
Total number of partners last year	1554	0.690	0.024	0.725	0.014	1.64	0.201
Number of non-spousal partners last year	1554	0.084	0.014	0.103	0.010	1.24	0.265
Condom use last sex	1096	0.052	0.014	0.088	0.010	4.57	0.033
Knows someone who has or died from AIDS	1531	0.551	0.025	0.625	0.014	6.37	0.012
Spoken with spouse about HIV	924	0.664	0.032	0.719	0.017	2.36	0.125
Willing to care for a relative with AIDS	1531	1.000	0.034	0.990	0.018	0.08	0.784
Teachers with AIDS should be allowed to teach	1530	0.877	0.064	0.862	0.036	0.04	0.838
Tested for HIV/AIDS	1526	0.136	0.018	0.135	0.010	0.00	0.975
People with AIDS are immoral	1531	1.102	0.116	0.911	0.057	2.17	0.141

Means are for women age 15-24 who do not live in Lilongwe district or at Likoma island. Standard errors are clustered at the district level for district inequality, district median wealth and Southern region, and at the cluster level for neighbourhood inequality and neighbourhood median wealth.

Paper II

J Popul Econ (2011) 24:629–655
DOI 10.1007/s00148-009-0303-2

ORIGINAL PAPER

Uncovering the impact of the HIV epidemic on fertility in Sub-Saharan Africa: the case of Malawi

Dick Durevall · Annika Lindskog

Received: 29 October 2008 / Accepted: 4 December 2009 /
Published online: 14 January 2010

Abstract We evaluate the impact of the HIV/AIDS epidemic on the reproductive behaviour for all women in Malawi, HIV-negative and HIV-positive alike, allowing for heterogeneous response depending on age and prior number of births. HIV/AIDS increases the probability that a young woman gives birth to her first child, while it decreases the probability to give birth of older women and of women who have already given birth. The resulting change in the distribution of fertility across age groups is likely to be more demographically and economically important than changes in the total number of children a woman gives birth to.

Keywords Africa · AIDS · Demographic transition

JEL Classification I10 · J13 · O12

1 Introduction

In many Sub-Saharan countries, HIV has spread to over 10% of the working-age population, making the epidemic a serious economic and social problem. There is no doubt that HIV/AIDS has wide-ranging consequences for households affected by the disease, but it is less obvious what the economic effects are at the national level. For instance, a number of studies have evaluated the nationwide economic effects of HIV/AIDS by testing its impact

Responsible editor: Junsen Zhang

D. Durevall · A. Lindskog (✉)
Department of Economics, University of Gothenburg,
P.O. Box 640, 405 30, Gothenburg, Sweden
e-mail: annika.lindskog@economics.gu.se

on growth, but with very mixed results.[1] One reason for the inconclusive results is probably that HIV/AIDS affects growth through many channels, whose relative importance is difficult to assess.

Recently, the effect of HIV/AIDS on fertility has emerged as one of the key channels. The course of fertility influences the future development of per capita GDP through its impact on the dependency ratio. In addition, changes in fertility and its distribution across ages affect various socioeconomic variables, such as child mortality, the number of orphans, and demand for education.

There is a large literature on the fertility of HIV-positive women, whose fertility is believed to decline for mainly physiological reasons (Fabiani et al. 2006). But there might also be a behavioural response to the HIV epidemic among women in general, due to changes in fertility preferences among both HIV-positive and HIV-negative women and changes in sexual behaviour to avoid infection. To date, there are only a few studies on behavioural changes among all women, and there is no consensus on the extent and direction of the effects. Young (2005, 2007) claims that women desire to have fewer children because of the HIV epidemic, which reduces fertility. Kalemli-Ozcan (2003, 2009) and Lorentzen et al. (2008) instead argue that theory suggests a positive association between child and adult mortality and fertility, i.e. that HIV-induced increases in mortality are likely to raise fertility. Analysing data on fertility in Sub-Saharan Africa, Young (2005, 2007) finds a large negative effect of HIV prevalence, while Kalemli-Ozcan (2009), as well as Juhn et al. (2008), Fortson (2009) and Fink and Linnemayr (2008) find the net effect on fertility to either be positive or non-existent.

An issue not addressed in the recent literature on HIV/AIDS and fertility is age-specific responses. Childbearing is an integral part of adulthood in most societies, and it is likely to be particularly important in societies where women have many children, such as in Malawi. This should make women respond differently depending on the stage of the reproductive life cycle. It is unlikely that women with no or few children would voluntarily abstain from giving birth; instead, they might give birth earlier when the probability of being HIV positive is lower. Moreover, women and men might also respond to the HIV epidemic by marrying and establishing supposedly monogamous relationships earlier, or men might try to marry younger women who are less likely to be HIV infected. These factors point to a difference in response between younger women, who have no or few children and a smaller probability of already being HIV infected, and older women, who probably have more children and a larger probability of being infected.

We evaluate the impact of communal HIV on realised fertility in Malawi, which is likely to be representative of at least its neighbouring countries. Our

[1] See, for instance, Bloom and Mahal (1997), Bell et al. (2004), Corrigan et al. (2005), Young (2005, 2007), McDonald and Roberts (2006), Werker et al. (2006), Papageoriou and Stoytcheva (2008) and Santaeulalia-Llopis (2008).

focus is on the behavioural response of the general female population, i.e. non-infected as well as infected women. In contrast to earlier studies, we include the possibility of a heterogeneous fertility response to the HIV epidemic, depending on the woman's age and number of children she has had.

Our main source of data is the Malawi Demographic and Health Survey (MDHS) carried out in 2000 and 2004, which includes the entire birth history of nationally representative samples of women. Using the retrospective birth information, we construct a panel of yearly observations from 1980 to the survey year for each woman. The birth history is then modelled as a discrete time process with a binary birth no-birth outcome, allowing for dependence on recent communal HIV prevalence rates, the woman's earlier birth history and other individual and communal characteristics.

Communal HIV prevalence rates have varied greatly over both time and space in Malawi. To measure this variation, we use district HIV rates obtained from women visiting antenatal clinics (ANCs). We have information from 18 of Malawi's 27 districts, covering about 75% of Malawi's population, which makes our measure of a communal effect different from other recent studies on HIV and fertility in Sub-Saharan Africa that either use countries (Young 2007; Kalemli-Ozcan 2009) or regions (Juhn et al. 2008; Kalemli-Ozcan 2009). There are three regions in Malawi, and HIV rates differ substantially within them.

To control for endogeneity of the spread of the HIV epidemic, we include district effects in our model. Year effects are also included to capture any fertility changes over time that might otherwise have been attributed to communal HIV/AIDS. Furthermore, we allow for unobserved individual effects. MDHS 2004 includes HIV status for a sub-sample of women. We use this information to verify that results are due to a behavioural response to the HIV epidemic, rather than being a biological difference in fertility between HIV-positive and HIV-negative women.

Our main finding is that the impact of HIV/AIDS on overall fertility is small and that the change in distribution of fertility across age groups is likely to be more important than any change in the total number of births per woman. As a response to the HIV epidemic, young women tend to give birth to their first child sooner; an increase in district HIV from 0% to 15%, would raise the probability of a first birth for women aged 20–24 by five percentage points, but the probability of another birth would fall by almost five percentage points for those who have at least three children. For women over 29 years, the probability of giving birth to a first, second, or third child, would decrease by seven to 15 percentage points, and the probability of women 35–39 giving birth to her fourth or fifth child would decrease by about five percentage points.

The next section discusses how HIV/AIDS might affect fertility in general, and reviews findings from previous studies. Section 3 gives a brief background to the HIV/AIDS epidemic and the evolution of fertility in Malawi. The empirical model and the data are described in Section 4, results are reported in Section 5, and their robustness is checked in Section 6. Section 7 discusses, summarizes and draws conclusions.

2 HIV/AIDS and fertility: theory and evidence

HIV/AIDS seems to affect fertility in numerous ways with no single theory explaining them all. Broadly speaking, it could affect fertility directly in two ways: biologically, because of the physiological consequences of the disease on fecundity, and behaviourally, because of changes in fertility preferences of both HIV-positive and HIV-negative women, as well as of changes in sexual behaviour to avoid infection.

Biologically, HIV seems to reduce fertility among infected women through a variety of mechanisms. The most important are believed to be: higher rates of miscarriage and stillbirth; co-infection with other sexually transmitted diseases; menstrual dysfunctions; weight loss leading to amenorrhoea; premature death of regular partner and reduced frequency of intercourse because of illness (Zaba and Gregson 1998; Fabiani et al. 2006). There is also an indirect impact on net fertility through premature deaths of women in their reproductive years (Lewis et al. 2004). The difference in fertility between HIV-positive and HIV-negative women has been shown by a number of studies (Gray et al. 1998; Zaba and Gregson 1998; Terceira et al. 2003; Fabiani et al. 2006). Except perhaps for girls aged 15–19, HIV-positive women have a lower fertility than HIV-negative women in all age groups, and the difference is considerable, 25–40% (Lewis et al. 2004).[2]

There is much less consensus on the role of behavioural mechanisms and how they impact fertility. Naturally enough, most research has focused on behavioural changes among HIV-positive women. But according to Fabiani et al. (2006), these are negligible, not least because few women know their HIV status (Gray et al. 1998; Zaba and Gregson 1998; Terceira et al. 2003). This seems to be the case in Malawi, where most people are not tested until they have already developed AIDS (Morah 2007). Moreover, a common finding in studies of women diagnosed with AIDS is that contraceptive use and fertility do not change much (Gray et al. 1998; Rutenberg et al. 2000; Oladapo et al. 2005), even though Yeatman (2009) finds that women who learn they are HIV positive desire fewer children than others.

Some recent research in economics has been concerned with changes in reproductive behaviour of women in general not only among HIV-positive women. Young (2005, 2007) argues that women respond to increased risk of HIV infection by having safe sex, and by reducing childbearing, to avoid getting infected and giving birth to HIV-infected babies. Grieser et al. (2001) find similar arguments in a qualitative study on Zimbabweans. Young (2005) also emphasizes another mechanism that could decrease fertility; theoretically high mortality rates among prime-age adults should reduce labour supply and increase wages, thus increasing the opportunity costs of children.

[2]The reason young HIV-positive women are more fertile is probably because they were more sexually active than uninfected women in the first place, and thus more likely to become both HIV-positive and pregnant. Since they became infected recently, they are still quite healthy (Ntozi 2002; Fabiani et al. 2006).

Yet, another behavioural mechanism that could lead to lower fertility relates to the increased number of orphans. In Malawi, nearly 13% of children aged 0–17 are orphaned, often due to AIDS (NSO and UNICEF 2008). The vast majority of these are taken care of by relatives, very few are in orphanages. Families that take care of orphans might decide to have fewer children of their own, both because of the costs and because children in their care might be substitutes for having their own children. In a qualitative study in Zambia, the extra burden of caring for AIDS orphans was the only reason given by symptomless women for curtailing future childbearing (Rutenberg et al. 2000). However, Zimbabweans told Grieser et al. (2001) that they could not depend on adopted children for old-age support and that taking care of orphans thus would not affect their own childbearing

There are also reasons why HIV/AIDS might increase fertility among women in general. Some young are reported to have expressed a desire for children soon, and grandparents and others are reported to have urged young women to have children while they are healthy, before they become HIV positive. And in countries with high HIV rates, such as Tanzania, Kenya and Uganda, total fertility rates seem to have increased or to have stopped decreasing recently (Measure DHS 2008). Estimates of the TFR in Malawi increased from 6.0 in 2004 (NSO and ORC Macro 2005) to 6.3 in 2006 (NSO and UNICEF 2008). Westoff and Cross (2006) attribute the increased TFR in Kenya to HIV/AIDS but do not test the hypothesis.

It has long been accepted that child mortality increases fertility, both through replacement and by creating a precautionary demand for children. And economic theory suggests that adult mortality would be positively related to fertility as well (Soares 2005; Lorentzen et al. 2008). Lower life expectancy in general leads to more children through the quantity–quality trade-off as returns to education decrease (Becker and Lewis 1973). Soares (2005) also suggests that there is substitutability between the number of children and the lifetime of each child directly in the parent's utility function and that parents care about the survival of their lineage which makes it important that enough children are alive and healthy through their reproductive period.

In a seminal paper, Young (2005) empirically tests the behavioural fertility effect of HIV/AIDS using data from South Africa. He finds that the HIV epidemic reduces fertility substantially, which will speed up the demographic transition and economic growth. In a later paper, Young (2007) obtains similar results using micro data from a sample of Sub-Saharan countries. Most other studies, however, do not confirm these results; they either find no effect – in most cases—or that the HIV epidemic increases fertility—in some cases.

Using macro data for 1960–2000, Lorentzen et al. (2008) find that adult mortality, which increases with HIV/AIDS, is positively associated with fertility. Soares (2006) also find that adult mortality increases fertility, using Brazilian micro data. Kalemli-Ozcan (2009) tests the association between HIV/AIDS and fertility directly, and her estimates based on between country variation shows that HIV/AIDS increases fertility. However, her within-country estimates show no effect on fertility. This latter result is confirmed by some other

studies; Werker et al. (2006) who use macro data with circumcision as an instrument to identify the causal impact of HIV/AIDS on fertility;[3] Magadi and Agwanda (2007), who study the effect of communal HIV rates on fertility in Kenya; Juhn et al. (2008), who use DHS data for 13 Sub-Saharan countries;[4] Fink and Linnemayr (2008) who use the latest DHS data, and World Fertility Surveys, from before the HIV epidemic, for five African countries;[5] and Fortson (2009) who use a panel of regional HIV and mortality rates and total fertility rates.

We argue that the fertility response to the HIV epidemic is likely to vary depending on age. The small and conflicting effects of the HIV epidemic on fertility could mask response heterogeneity across age groups. There are a few studies that point in this direction: In a qualitative study, Grieser et al. (2001) find evidence that some women wish to have children quickly, before they become HIV positive, sometimes seeing this as unavoidable; Ueyama and Yamauchi (2009), using MDHS 2004 data, find that Malawian women marry earlier if district adult mortality is higher; and Noël-Miller (2003), using Malawian data from another survey, finds that the association between the degree of anxiety about becoming HIV infected and the number of births is positive among young women but negative among older women.

3 HIV/AIDS and fertility in Malawi

Malawi's first AIDS case was diagnosed in 1985, at a time when the national HIV prevalence rate was very low. From then and on, the HIV epidemic spread rapidly, first in the major cities, and then in rural areas. By the mid-1990s, HIV prevalence reached 14%, where it stabilized. In urban areas, it peaked at close to 25% in 1995 and then started to decline slowly. Current overall HIV prevalence rates are 15.6% in urban areas and 11.2% in rural areas (GoM 2007). The actual variation is much greater, however; from 5% in some of Malawi's 27 districts to over 20% in others (NSO and ORC Macro 2005).

We know people had practically no knowledge of HIV/AIDS in the early 1980s. By now, the disease has been around in Malawi for over 25 years, raising prime-age adult mortality about four times; three out of four deaths among those aged 15–49 are due to AIDS (Doctor and Weinreb 2003). As a result, knowledge about HIV/AIDS is widespread. In fact, already in the MDHSs carried out in 1992, about 90% of the respondents had heard about the disease,

[3] Randomised experiments have shown that circumcision reduces the risk of becoming infected with HIV.

[4] Juhn et al. (2008) find no effect among HIV negative women, but, in line with the literature referred to in the beginning of this section, HIV positive women have 20% lower fertility than HIV negative women.

[5] The countries studied by Fink and Linnemayr (2008) are Cameron, Cote d'Ivoire, Ghana, Kenya and Senegal.

and by 2000, the number had risen to 99%. Moreover, according to the 1992 MDHS, 86.3% of the women and 93.2% of the men who had heard about AIDS knew that it was transmitted through sexual intercourse.

Table 1 provides data on some additional indicators of HIV/AIDS knowledge and use of contraceptives from the 1992 and 2004 MDHSs. In the beginning of the 1990s, 26% of women and 11.6% of men did not think AIDS could be avoided; nowadays, practically all know this is wrong. The fact that less than 30% believed that using condom or abstaining from sex could prevent infection, according to the 1992 MDHS, shows that there was widespread ignorance at that time. The 2004 DHS reports much higher percentages, albeit still low in some cases; only 56% of the women thought condoms work. There is also evidence of a decrease in the belief of limiting sex to one partner as an HIV-prevention method, which might appear strange. However, it is probably an indication of increased awareness that the partner might be HIV-positive.

In the early 1990s, somewhat over 80% knew that the HIV virus can be transmitted from mother to child; the knowledge was about the same in the 2004 MDHS. The widespread knowledge of mother-to-child transmission is one potential mechanism through which HIV/AIDS might affect fertility, since it creates fear of giving birth to HIV-infected babies. In fact, the data on contraceptive use indicates an increased concern for controlling fertility; among married couples, as many as 32.5% currently used some type of contraception, and 28.1% used a modern method, such as the pill or injectables, according to the 2004 MDHS, compared to 13.0% and 7.4% in the early 1990s. The use of modern methods had thus tripled. This could be part of a general trend towards lower demand for children and increased access to contraceptives but is also consistent with HIV/AIDS-induced reductions in fertility, as described by Young (2005, 2007) In any case, the overall picture is that awareness of HIV and AIDS has gone from being non-existent in the 1980s to be quite general in the 2000s. Hence, if HIV/AIDS affects decision-making about childbearing, this should be visible in Malawi.

Table 1 Indicators of HIV knowledge and use of contraceptives

	1992 DHS		2004 DHS	
	Men	Women	Men	Women
Believe there is no way to avoid AIDS	11.6	26.2	0.8	2.6
Knowledge of prevention methods				
Using condoms	27.2	8.2	71.0	56.3
Limiting sex to one partner	49.0	28.2	18.7	23.3
Abstaining from sex	32.1	25.0	88.4	75.0
Can AIDS be transmitted from mother to child				
Yes	87.6	81.6	81.1	81.2
No or do not know	12.2	18.3	18.9	18.9
Current use of contraceptives %				
Any method		13.0[a]		32.5[a]
Any modern method		7.4[a]		28.1[a]

[a]Married couples
Source: (NSO and ORC Macro 2001, 2005)

In the early 1960s, the fertility rates in Malawi were similar to those in other developing countries. But while fertility in most other countries fell, Malawi's grew until 1980, probably because of the ideology and policy of the Malawian government under President Banda: birth control was seen as incompatible with Malawian culture (Chimbwete et al. 2005). Finally, in the beginning of the 1980s fertility started to fall as it had in many other developing countries. Nonetheless, it is still higher than in most African countries. In the 2000s, the TFR has been 6.0 to 6.3, implying that women in Malawi on average give birth to one child more than the average African woman (NSO and ORC Macro 2005; NSO and UNICEF 2008).

There are substantial differences in how age-specific fertility has evolved over time (see Fig. 1). Fertility among women over 30 fell greatly during the 1990s, accounting for most of the downward trend during this period, while it only fell slightly among 20–24- and 25–29-year-olds and actually grew slightly among 15–19-year-olds. From 2000 to 2006, a period when HIV/AIDS knowledge was widespread and practically everybody had personal experiences of the epidemic, fertility appears to have remained mostly stable for older women, while still showing some increases among younger women. The dearth of a fertility decrease among young women is surprising since the years of schooling has increased over the period, and it is generally believed that increased female education decreases fertility.

The observation that fertility fell among older women during the 1990s, while it rose among younger women, is consistent with a heterogeneous response to the HIV epidemic across age groups, with younger women seeking to

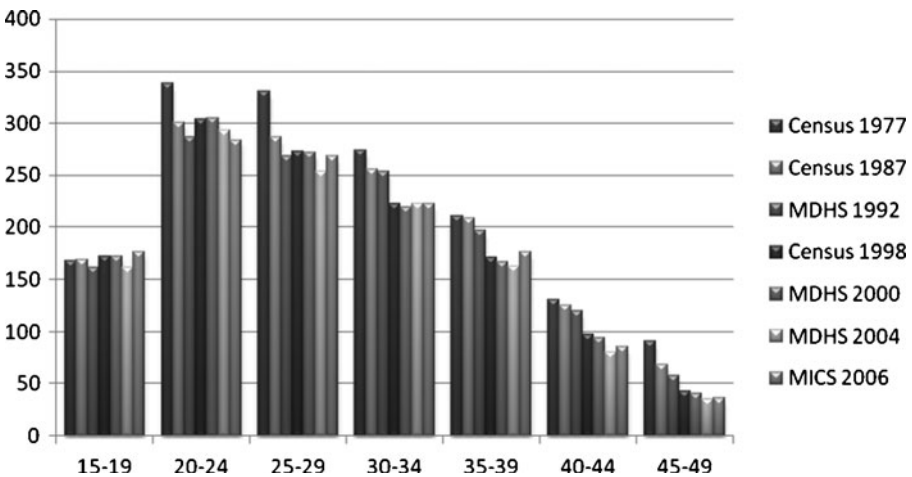


Fig. 1 Age-specific fertility rates (births per 1,000 women) for various years. Population Census 1977, 1987 and 1998 (NSO 1993, 2002), MDHS 1992, 2000 and 2004 (NSO and ORC Macro 2001, 2005) and MICS 2006 (NSO and UNICEF 2008)

give birth earlier, and with older women, who already have children, reducing additional births.

4 Empirical framework and data

This section first outlines and explains the empirical model, then describes the data and variables in some detail.

4.1 Empirical model

When analysing the impact of the HIV epidemic on fertility we are ultimately interested in the effect on a woman's complete birth history, the total number of children she gives birth to and the timing of those births. However, there are still no women whose whole reproductive life cycle has taken place during the HIV epidemic, which makes it impossible to study complete birth histories directly. Following Barmby and Cigno (1990), a suitable approach is thus to treat childbearing as a sequential choice where decisions are taken under uncertainty. The birth history is then modelled as a discrete time process with a binary birth no-birth outcome in each period. This approach allows fertility choice to depend on the most recent information on the HIV epidemic during each period, as well as on earlier birth history and other individual and communal characteristics. Even though we do not study complete birth histories, we can, within one coherent model, study the fertility of women at different stages in their reproductive life cycles (at different ages and with different number of earlier births), while allowing their response to the HIV epidemic to differ.

To illustrate the approach we outline a simple version of the model. Let $Y_{it} = 1$ if woman i gives birth in year t, and $Y_{it} = 0$ if she does not. A woman decides whether to have a child or not during a given year by taking into account a number of individual and communal factors, particularly HIV rates, though of course, she cannot control the exact timing and number of births for both biological and behavioural reasons. Formally,

$$\begin{array}{ll} Y_{it} = 1 & \text{if} \quad X'_{it}\beta + Z'_{it}\text{HIV}'_{dt-1}\gamma + \varepsilon_{it} > 0 \\ Y_{it} = 0 & \text{otherwise} \end{array} \tag{1}$$

where β and γ are coefficients, X_{it} indicates individual characteristics of the woman; HIV_{dt-1} is last year's district HIV prevalence rate; and Z_{it}, which is a sub-set of X_{it}, indicate the woman's age and her number of previous births. With the exception of the interaction term, $Z'_{it}\text{HIV}'_{dt-1}$, this is a standard binary model. By letting HIV rates enter the model through the interaction terms in $Z'_{idt}\text{HIV}'_{dt-1}$ rather than independently, we allow for a differential impact of the district HIV rate depending on the factors in Z_{it}. The variables in Z_{it} are dummies, and the interaction term effects should thus be interpreted as the effect of HIV on fertility in the particular group. The impossibility to exactly control fertility gives rise to an error term, ε_{it}. In practice, the error

term will also capture factors unobserved by the researcher. Assuming ε_{it} to be logistically distributed, we can use the logit estimator.

We have so far not taken into account the fact that the spread of the HIV epidemic across time and space is not exogenous. Norms of sexual and reproductive behaviour and other factors affecting them will have an impact on both fertility and the spread of HIV. We thus add district effects, α_d, to control for district-level unobserved heterogeneity. Moreover, the number of people infected with HIV, and dying of AIDS, has increased over time. To capture unobserved time-varying effects, we, therefore, add year effects δ_t. With a small number of districts and years, and many observations per district and per year, as in our case,[6] we can estimate the logit model with dummy variables to capture the district and year effects, obtaining consistent parameter estimates. We, therefore, use this simple dummy variable approach. Estimations are carried out maximizing the (standard logit) likelihood function

$$L = \prod_{it} \left[\Delta\left(X'_{it}\beta + Z'_{it}\mathrm{HIV}'_{dt-1}\gamma + \delta_t + \alpha_d\right)\right]^{Y_{it}} \times \left[1 - \Delta\left(X'_{it}\beta + Z'_{it}\mathrm{HIV}'_{dt-1}\gamma + \delta_t + \alpha_d\right)\right]^{1-Y_{it}} \quad (2)$$

where $\Delta(.)$ denotes the logistic cumulative distribution function. To draw inference, we use standard errors clustered at the level of the DHS sampling clusters.[7]

In the robustness analysis, we also allow for unobserved heterogeneity at the level of the individual, i.e. that women with the same observable characteristics might behave differently. To do so, we divide the error term (ε_{id}) into a time-constant individual unobserved effect, u_i, and time-varying component, v_{it}. Given the large number of women and the small number of observations per woman, the simple dummy-variable approach would yield biased estimates due to the incidental parameters problem first presented in Neyman and Scott (1948). Instead, a random effects estimator is used. The likelihood function is then written as the likelihood of a certain sequence of birth outcomes for each woman marginal on the random effects, making it a function of the parameters β, γ, δ, α and the parameters describing the distribution of the random effects but, importantly, not of each u_i. Let L_i be the contribution of woman i to the likelihood function, which is the joint probability of all her T_i observations. This is an integral of dimension T_i, computationally demanding to solve. To ease the computational burden, unobserved group effects, here

[6] In the main estimations, we have almost 150,000 observations (in the smallest sample in the robustness analysis almost 15,000 observations) distributed over 25 years and 18 districts.

[7] Clustering the standard errors at the district level does not significantly change our results, probably because district dummies capture most of the district correlation. Estimations with standard errors clustered by district are available from the authors.

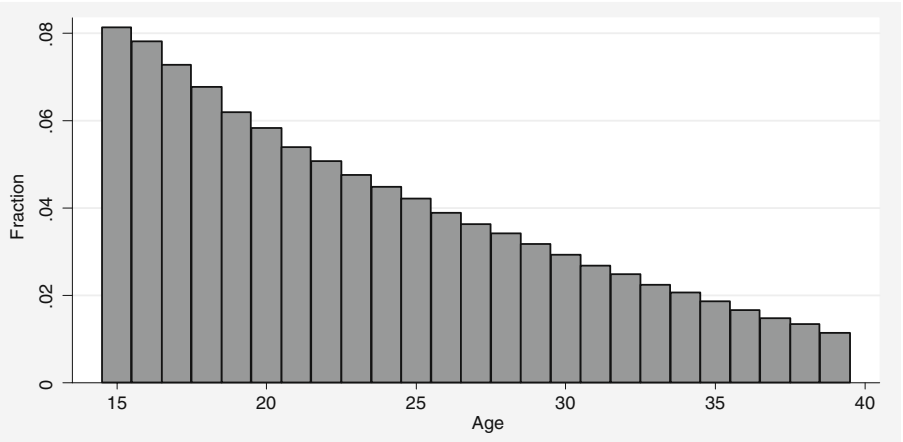


Fig. 2 Age distribution of the estimation sample. Own calculations using data from Malawi Demographic and Health Surveys 2000 and 2004

u_i, are integrated out of the likelihood function,[8] which gives a conditional likelihood (on unobserved group effects) of the form

$$L_i = \int_{-\infty}^{\infty} \left[\prod_t \Pr\left(Y_{it} = 1 \,|\, X_{it}, \mathrm{HIV}_{dt-1}, u_i\right)^{Y_{it}} \times \left(1 - \Pr\left(Y_{it} = 0 \,|\, X_{it}, \mathrm{HIV}_{dt-1}, u_i\right)\right)^{1-Y_{it}} \right] f(u_i)\, du_i \qquad (3)$$

where the term inside the bracket in our case is the logit model. The solution now requires only one-dimensional integration. However, it is, necessary to make an assumption about the random effects distribution to solve the equation. It is common that the unobserved group effects are assumed to be normally distributed, and integration is done using numerical methods. This approach is also used here.[9] Again, standard errors are clustered at the level of the sampling cluster.

4.2 Data and variables

The data used is from MDHS 2000 and 2004.[10] In MDHS 2000, 13,220 women were interviewed, and in MDHS 2004, 11,698 women. The Demographic and Health Surveys project collects information about the entire birth history of

[8] Conditional on unobserved group effects, the error terms ε_{it} are independent, and their joint probability is consequently equal to the product of the probability of each term.

[9] Maximization is done by adaptive Gaussian quadrature using the gllamm procedure in Stata.

[10] Available at http://www.measuredhs.com/.

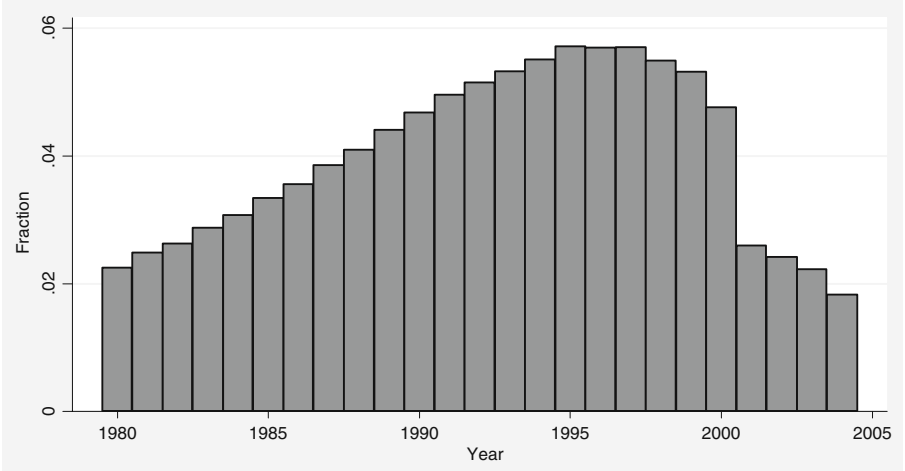


Fig. 3 The distribution per year of the estimation sample. See Fig. 2

interviewed women. Using this retrospective information, we create a panel dataset consisting of one observation for each woman and year. A woman enters the sample at age 15, or in 1980, if she was older than that then. She leaves the sample when she turns 40, or earlier if she is still not 40 in the survey year. The choice of 1980 as the start year is a compromise between the desire to include observations from before the onset of the HIV epidemic and the desire to include women from different age groups over time. This is also the reason to exclude women that are over 39 from the sample. Still, both the year distribution and the age distribution of the data are skewed with more observations for younger women and for later years (Figs. 2 and 3). There are especially few observations for older women in the 1980s.[11] However, the skewed distribution should not be a problem since we use time and age dummies.

In total, we have 296,067 woman-year observations for 24,915 women. However, we only use observations when we know where the woman lived during a particular year, where we have at least two observations for the woman, and from districts with HIV prevalence data, which reduces the number of woman-year observations to 148,166 for 14,241 women.[12]

[11]Women were sampled to be representative of 15–49-year-old women in the survey years, resulting in a younger age-distribution as we go back in time.

[12]The sample was further reduced because of missing information regarding ethnicity for six women and regarding relative household wealth for another six women.

Table 2 Distribution of prior births for different age groups

	Age 15–19	Age 20–24	Age 25–29	Age 30–34	Age 35–39
No prior births	0.82 [0.12]	0.27 [0.26]	0.07 [0.19]	0.03 [0.11]	0.02 [0.04]
One or two prior births	0.17 [0.23]	0.56 [0.31]	0.31 [0.30]	0.15 [0.19]	0.09 [0.09]
Three or four prior births	0.01 [0.21]	0.17 [0.24]	0.46 [0.28]	0.33 [0.26]	0.20 [0.17]
Five or more prior births	0.00 [0.25]	0.01 [0.22]	0.16 [0.24]	0.49 [0.25]	0.68 [0.20]
Total	1.000 [0.14]	1.000 [0.28]	1.000 [0.28]	1.000 [0.24]	1.000 [0.18]

Proportion of positive birth outcomes in brackets; see Fig. 2

As mentioned earlier, the dependent variable is binary, equalling 1 if the woman gave birth during a particular year, 0 otherwise.[13] A birth was recorded in about a fifth of the cases, 32,211 observations. Table 2 shows the distribution of number of prior births for different age groups and the proportion (in brackets) who gives birth in each category. There are very few observations for 15–19-year-olds with three or more prior births and for 20–24-year-olds with five or more births. And, as shown in the brackets, among the few women over 29 who have not given birth earlier, very few do it later. Still, it is clear that women in Malawi give birth to many children and start childbearing early. The proportion of no-prior-births decreases rapidly with age, from 82% for those aged 15–19 to 27% for 20–24, and then 7%, 3% and 2% for 25–29, 30–34 and 35–39. By summing the proportions, we can also see that over 70% of women aged 20–24 have given birth at least once before, over 60% of women aged 25–29 and over 80% of women aged 30–34 have given birth at least three times and almost 70% of women aged 35–39 have given birth to as many as five children or more.

We are interested in the effects of the HIV epidemic on the fertility behaviour of all women, not just the HIV-positive. For this purpose, we need a variable measuring the geographic and time variation in risk of HIV infection. We use district-level HIV prevalence rates collected from pregnant women visiting ANCs.[14] The raw data consist of observations for selected years from 1985 to 2003 for a maximum of 18 clinics across Malawi. It covers about 75% of Malawi's population, with a bias towards urban dwellers since all districts with major cities are included. This is the only data on HIV prevalence rates that has been collected reasonably systematically over a longer time period.[15]

Worries are often expressed over how well the ANC data represent HIV prevalence rates in the general population. However, our statistical identification of the HIV impact comes from relative levels of HIV prevalence rates over time and space. Thus, the HIV rates measured at the ANC do not have to

[13]The few cases where a woman gave birth to more than one child in a year are thereby not treated differently than cases where a woman gave birth to one child.

[14]The data is provided by the US Census Bureau in the HIV surveillance database, http://www.census.gov.

[15]In a related study, Durevall and Lindskog (2009), we use district adult mortality in 1998 and HIV prevalence in the general population in 2004. Though these variables might be of better quality, they only allow for cross-sectional analysis. The results are, however, very similar.

be correct, as long as any bias is similar across clinics and over time. The use of HIV prevalence rates in empirical studies of human behaviour has also been questioned since they are not directly observable by people. But they might be indicative of observable AIDS illness and deaths. Young (2007) argues that women are able to infer the HIV rate in their community from infants' deaths with AIDS symptoms, since the disease progresses rapidly in small children.

To obtain HIV prevalence rates for the 18 districts included in the analysis, and for all years from 1980 to 2004, we used the Estimation and Projections Package (EPP) of WHO/UNAIDS to estimate HIV trends. The EPP is constructed specifically for ante-natal surveillance data from countries with high HIV prevalence rates. It fits a (nonlinear) epidemiological model to the data. We set prevalence rates to zero in 1980, assuming there are hardly any known AIDS cases, i.e. people do not worry about the disease.[16]

When modelling fertility as a sequential choice, it is crucial to control for the woman's age and number of earlier births. Age enters as five dummy variables for age groups, as seen in Table 2: 15–19 (the baseline in estimations); 20–24; 25–29; 30–34; and 35–39. Dummies indicating how many previous births the woman has had are also included; no prior births (the baseline in estimations); one or two prior births; three or four prior births and five or more prior births. Since one purpose of this study is to allow the fertility response to the HIV epidemic to be conditional on age and number of previous births, we construct various interaction terms with HIV rates and the age and birth dummies.

According to economic theory, family income as well as the opportunity cost of having a child should influence fertility (Becker and Lewis 1973). As a proxy for family income, dummies for the wealth quintile of the household are used (the middle quintile dummy is the baseline in estimations).[17] To capture the opportunity cost, the woman's education is included, indicated by dummies for no or incomplete primary school (the baseline in estimations); completed only primary school and completed secondary school. A problem with the education and wealth variables is that we have information from the survey year only. We thus have to assume that there have not been any systematic changes in the relative wealth of the households over time, and we assume that primary school was completed by age 15 and secondary school by age 20. These assumptions are checked in the robustness analysis section by estimating the models without the education and wealth variables.

Norms and social constraints are also important determinants of fertility. They are likely to differ depending not only on education and wealth but also on ethnicity. Through its influence on sexual and reproductive behaviour, ethnicity should matter both for fertility and the spread of HIV/AIDS. Nine dummies indicating ethnic group are thus included to capture variation in

[16] Young (2007) and Kalemli-Ozcan (2009) also used the Estimation and Projections Package (EPP) to create time series for HIV rates.

[17] The household wealth variable has been constructed using information on household assets. See Rutstein and Johnson (2004) for further information about the DHS wealth index.

Table 3 Summary statistics

Variable	Obs	Mean	SD	Min	Max
Gave birth last 12 months	148,020	0.217	0.412	0.000	1.000
Age 15–19	148,020	0.358	0.479	0.000	1.000
Age 20–24	148,020	0.257	0.437	0.000	1.000
Age 25–29	148,020	0.184	0.388	0.000	1.000
Age 30–34	148,020	0.125	0.331	0.000	1.000
Age 35–40	148,020	0.076	0.264	0.000	1.000
No prior births	148,020	0.380	0.486	0.000	1.000
One or two prior births	148,020	0.289	0.453	0.000	1.000
Three or four prior births	148,020	0.186	0.389	0.000	1.000
Five or more prior births	148,020	0.145	0.352	0.000	1.000
Poorest household wealth quintile	14,230	0.170	0.376	0.000	1.000
Second poorest household wealth quintile	14,230	0.187	0.390	0.000	1.000
Middle household wealth quintile	14,230	0.200	0.400	0.000	1.000
Second richest household wealth quintile	14,230	0.206	0.404	0.000	1.000
Richest household wealth quintile	14,230	0.236	0.425	0.000	1.000
Child mortality among siblings	14,230	0.151	0.214	0.000	1.000
Number of siblings	14,230	5.760	2.733	0.000	18.000
Urban residence	14,230	0.205	0.372	0.000	1.000
Completed primary school, only	14,230	0.685	0.440	0.000	1.000
Completed secondary school	148,020	0.054	0.226	0.000	1.000
Last year's district HIV rate	701	10.250	9.086	0.000	32.390
Chewa	14,230	0.303	0.460	0.000	1.000
Tumbuka	14,230	0.127	0.333	0.000	1.000
Lomwe	14,230	0.144	0.351	0.000	1.000
Tonga	14,230	0.026	0.158	0.000	1.000
Yao	14,230	0.157	0.364	0.000	1.000
Sena	14,230	0.029	0.168	0.000	1.000
Nkonde	14,230	0.028	0.165	0.000	1.000
Ngoni	14,230	0.123	0.328	0.000	1.000
Other ethnicity	14,230	0.045	0.207	0.000	1.000
Blantyre	14,230	0.097	0.296	0.000	1.000
Kasungu	14,230	0.091	0.287	0.000	1.000
machinga	14,230	0.089	0.284	0.000	1.000
Mangochi	14,230	0.086	0.280	0.000	1.000
Mzimba	14,230	0.093	0.291	0.000	1.000
Lilongwe	14,230	0.092	0.289	0.000	1.000
Mulanje	14,230	0.102	0.303	0.000	1.000
Karonga	14,230	0.064	0.244	0.000	1.000
Nkhata bay	14,230	0.018	0.132	0.000	1.000
Rumphi	14,230	0.012	0.107	0.000	1.000
Nkhota kota	14,230	0.022	0.147	0.000	1.000
Dowa	14,230	0.050	0.219	0.000	1.000
Mchinji	14,230	0.037	0.190	0.000	1.000
Dedza	14,230	0.055	0.229	0.000	1.000
Ntcheu	14,230	0.048	0.214	0.000	1.000
Chiradzulu	14,230	0.023	0.149	0.000	1.000
Nsanje	14,230	0.020	0.140	0.000	1.000
Mzuzu city	14,230	0.001	0.027	0.000	1.000

Summary statistics are based on all observations for time varying individual variables, one observation per woman for time constant individual variables, and one observation per district and year for 'Last year's district HIV rate'

norms and social constraints. Urban or rural residence is also controlled for with an urban-residence dummy, which is probably associated with both differences in opportunity costs of having children and with differences in norms.

Following Soares (2006), we also allow for intergenerational persistence in reproductive behaviour by including the woman's number of siblings, and the effects of child mortality in the woman's original family, as measured by the share of her siblings that died before age 10. Table 3 shows summery information of the main variables.

5 Empirical results

We estimate four specifications of our fertility model, with the degree of heterogeneity allowed for in the fertility response to communal HIV differing across them (Table 4). In the first specification, the fertility response to communal HIV is constrained to be equal for all women. In the second specification, it is allowed to differ across age groups and in the third across prior number of births. In the fourth and most flexible specification, the response is allowed to differ for each combination of age and prior number of births.

In all four specifications, women older than 19 have a higher probability of giving birth than women aged 15 to 19, with child-bearing peaking in the early 20s. Women who have already had at least one child have a higher probability of giving birth than those with no prior children. This is in line with the findings of, for example, Barmby and Cigno (1990) and Angeles et al. (2005). Given at least one prior birth, the probability of giving birth again seems to be decreasing with number of earlier births, though there are small differences between 'three or four' and 'five or more' births.

Women in richer households give birth to fewer children than women in poor and middle-wealth households, which could suggest that the substitution effect is larger than the income effect. Women with more education give birth to fewer children, especially those few who have completed secondary school. This is in line with the hypothesis of lower fertility as the opportunity cost of having children rises, but it could also be because better informed women, or women with a better bargaining position in the household, decide to have fewer children. As expected, women in cities have fewer children, which could be because of the higher opportunity costs of having children in cities, because of a larger need for (child) labour among farmers or perhaps because of differences in norms. As in the analysis on Brazil by Soares (2006), Malawian women who had more siblings give birth to more children, as do those whose siblings died as children.[18]

[18]The coefficient on number of siblings, that is the number of children the mother of the woman has given birth to, is almost statistically significant at the 10% level in all specifications.

So much for the control variables. In the first specification, the estimated coefficient of HIV prevalence is negative, as in Young (2007), but not statistically significant.[19] The lack of significance could be because of heterogeneity in the fertility response to the HIV epidemic. Results from the remaining specifications support this. In the second specification, there is a statistically significant negative effect of HIV rates on fertility for women over 24, particularly for women over 29. In the third specification, the probability of a first birth is positively related to HIV rates, while the probability of subsequent births is negatively related to HIV. In the fourth and most flexible specification, higher HIV rates increase the probability of a first birth for women aged 20–24—the age at which fertility is highest for Malawian women—while there is a smaller and imprecisely measured change for women 15–19 and a decrease in the probability for women over 29. For women who have already had at least one child, the probability of another birth falls with higher HIV rates, with the exception for women aged 15–19 who have a higher probability of another birth.

To formally compare the four specifications, we re-estimated them without clustered standard errors and performed log-likelihood ratio tests. The null hypothesis that the fertility response to HIV is homogenous—first specification—is strongly rejected at the 1% critical level in favour of age or birth interval specific fertility responses—second and third specifications ($\chi^2(4) = 38.9$ and $\chi^2(3) = 55.7$). Similarly, the null hypothesis of the HIV response depending on age but being homogenous within age groups—second specification—was strongly rejected at the 1% critical level in favour of the HIV response being age and birth-interval-specific—fourth specification ($\chi^2(15) = 501.8$), as was the null hypothesis of the HIV response depending on birth interval but being homogenous within birth intervals—the third specification ($\chi^2(16) = 485$).

These results support the hypothesis that young women start child-bearing earlier when HIV rates are high, while older women are less likely to give birth even if they have no children. The results come out even though we control for the district in which the woman lives and her ethnicity, which means that results probably are not due to any initial differences in sexual and reproductive behaviour across districts or ethnic groups.

To illustrate the magnitude of the effects of HIV on fertility, predicted probabilities of women giving birth with a district HIV rate at 0% and 15% are calculated (Table 5). Among those aged 15 to 19, HIV rates do not alter much; the difference in the probability of a birth is an increase of around one percentage point, which is not significant at the 5% level. But for women aged 20–24, the probability of a first birth is five percentage points higher when HIV rates are 15% (clearly significant at the 5% level), whereas the probability of

[19]There are several potential explanations for differences in results even though Young (2007) also use individual level data from the DHS project. For example, our measure of the communal effect is based on much smaller geographical areas, districts, whereas Young (2007) uses countries. Moreover, we use annual time dummies, while Young (2007) includes a linear time trend in some specifications.

Table 4 Logit estimates of the fertility effect of the HIV epidemic

	(1)	(2)	(3)	(4)
Age 20–24	0.691[c]	0.716[c]	0.692[c]	0.671[c]
	(0.023)	(0.029)	(0.023)	(0.031)
Age 25–29	0.613[c]	0.679[c]	0.616[c]	0.641[c]
	(0.028)	(0.034)	(0.028)	(0.038)
Age 30–34	0.434[c]	0.538[c]	0.441[c]	0.528[c]
	(0.034)	(0.041)	(0.034)	(0.047)
Age 35–39	0.098[b]	0.327[c]	0.109[c]	0.345[c]
	(0.042)	(0.057)	(0.042)	(0.062)
One or two prior births	0.492[c]	0.492[c]	0.615[c]	0.611[c]
	(0.019)	(0.019)	(0.026)	(0.028)
Three or four prior births	0.329[c]	0.330[c]	0.445[c]	0.408[c]
	(0.025)	(0.025)	(0.032)	(0.036)
Five or more prior births	0.360[c]	0.358[c]	0.445[c]	0.346[c]
	(0.031)	(0.031)	(0.040)	(0.046)
Poorest quintile	0.003	0.001	0.002	−0.003
	(0.020)	(0.020)	(0.020)	(0.020)
Second poorest quintile	0.011	0.011	0.011	0.013
	(0.020)	(0.020)	(0.020)	(0.020)
Second richest quintile	−0.062[c]	−0.062[c]	−0.061[c]	−0.061[c]
	(0.020)	(0.020)	(0.020)	(0.020)
Richest quintile	−0.101[c]	−0.102[c]	−0.103[c]	−0.101[c]
	(0.025)	(0.025)	(0.025)	(0.025)
Child mortality among siblings	0.087[c]	0.083[c]	0.083[c]	0.084[c]
	(0.029)	(0.030)	(0.030)	(0.029)
Number of siblings	0.003	0.004	0.004	0.004
	(0.002)	(0.002)	(0.002)	(0.002)
Urban residence	−0.071[c]	−0.074[c]	−0.074[c]	−0.073[c]
	(0.026)	(0.026)	(0.026)	(0.026)
Completed primary school	−0.064[c]	−0.067[c]	−0.066[c]	−0.071[c]
	(0.015)	(0.015)	(0.015)	(0.015)
Completed secondary school	−0.337[c]	−0.337[c]	−0.341[c]	−0.397[c]
	(0.035)	(0.035)	(0.035)	(0.036)
District HIV prevalence	−0.002			
	(0.001)			
District HIV prevalence interaction terms				
Age 15–19		0.003		
		(0.002)		
Age 20–24		0.000		
		(0.002)		
Age 25–29		−0.003[a]		
		(0.002)		
Age 30–34		−0.006[c]		
		(0.002)		
Age 35–39		−0.014[c]		
		(0.003)		
No prior births			0.005[c]	
			(0.002)	
One or two prior births			−0.006[c]	
			(0.002)	
Three or four prior births			−0.005[c]	
			(0.002)	
Five or more prior births			−0.003	
			(0.002)	
No prior births age 15–19[a]				0.002
				(0.002)

Table 4 (continued)

	(1)	(2)	(3)	(4)
No prior births age 20–24[a]				0.020[c]
				(0.003)
No prior births age 25–29[a]				−0.005
				(0.005)
No prior births age 30–34[a]				−0.043[c]
				(0.011)
No prior births age 35–39[a]				−0.107[c]
				(0.033)
One or two prior births age 15–19[a]				0.004[a]
				(0.002)
One or two prior births age 20–24[a]				−0.004[b]
				(0.002)
One or two prior births age 25–29[a]				−0.004
				(0.002)
One or two prior births age 30–34[a]				−0.036[c]
				(0.004)
One or two prior births age 35–39[a]				−0.071[c]
				(0.009)
Three or four prior births age 15–19[a]				0.007
				(0.009)
Three or four prior births age 20–24[a]				−0.016[c]
				(0.003)
Three or four prior births age 25–29[a]				−0.001
				(0.002)
Three or four prior births age 30–34[a]				0.000
				(0.003)
Three or four prior births age 35–39[a]				−0.022[c]
				(0.004)
Five or more prior births age 15–19[a]				−0.832[c]
				(0.230)
Five or more prior births age 20–24[a]				−0.019
				(0.013)
Five or more prior births age 25–29[a]				−0.012[c]
				(0.003)
Five or more prior births age 30–34[a]				−0.000
				(0.002)
Five or more prior births age 35–39[a]				−0.005
				(0.003)
Observations	148,020	148,020	148,020	148,020
Log pseudo likelihood	−74,735.89	−74,716.39	−74,708.03	−74,465.53
Pseudo R^2	0.0354	0.0356	0.0357	0.0388

Dependent variable is birth/no birth; all estimations include a constant, ethnicity dummies, year dummies and district dummies. Robust standard errors, clustered at the level of the sampling cluster, are in parentheses
[a]Significant at 10%
[b]Significant at 5%
[c]Significant at 1%

another birth is almost five percentage points lower for women that have at least three children already (significant at the 5% level for women with three or four prior births, but not for the smaller number of women with five or more prior births). There are also large effects among women over 29, whose

Table 5 Predicted probabilities of giving birth during any given year

	No HIV	15% HIV prevalence rate	Difference	95% confidence interval of difference
No prior births				
Age 15–19	0.118	0.121	0.003	[−0.003, 0.010]
Age 20–24	0.195	0.246	0.051	[0.0380, 0.064]
Age 25–29	0.198	0.188	−0.011	[−0.031, 0.010]
Age 30–34	0.185	0.107	−0.079	[−0.110, −0.047]
Age 35–39	0.158	0.037	−0.122	[−0.159, −0.084]
One or two prior births				
Age 15–19	0.202	0.212	0.010	[−0.002, 0.022]
Age 20–24	0.323	0.310	−0.012	[−0.025, −0.000]
Age 25–29	0.312	0.301	−0.011	[−0.026, 0.004]
Age 30–34	0.290	0.193	−0.098	[−0.119, −0.076]
Age 35–39	0.251	0.103	−0.148	[−0.178, −0.117]
Three or four prior births				
Age 15–19	0.176	0.192	0.016	[−0.027, 0.059]
Age 20–24	0.289	0.242	−0.047	[−0.063, −0.032]
Age 25–29	0.279	0.275	−0.003	[−0.016, 0.009]
Age 30–34	0.253	0.253	0.000	[−0.015, 0.016]
Age 35–39	0.215	0.164	−0.051	[−0.071, −0.031]
Five or more prior births				
Age 20–24	0.287	0.232	−0.055	[−0.122, 0.011]
Age 25–29	0.273	0.238	−0.035	[−0.053, −0.017]
Age 30–34	0.248	0.247	−0.001	[−0.014, 0.012]
Age 35–39	0.212	0.200	−0.011	[−0.026, 0.004]

Predicted probabilities are based on parameter estimates in specification 4 in Table 4. The value of characteristics other than age, prior births, and district HIV prevalence are set to the mean for women in each age and prior births group. Predicted probabilities for 15–19-year-olds with five or more births are not reported as there are only four observations and one woman in this category

probability of giving birth to a first, second, or third child, are eight to 15 percentage points lower with a 15% HIV rate, the differences being clearly significant at the 5% level. The probability of women 35–39 giving birth to her fourth or fifth child is about five percentage points lower and significant at the 5% level. It thus seems that the HIV epidemic changes the distribution of fertility across age groups, leading to more births for younger women and fewer for older women.

Using the predicted probabilities in Table 5, we calculated the predicted probability to give birth during a year for a women representative of Malawi's 15–39-year-olds with no HIV and a 15% prevalence rate The age distribution of women in the 1998 population census (NSO 2002) and the distribution of total number of births within age groups in the MDHS 2004 final report (NSO and ORC Macro 2005) were used to provide weights. The impact of the HIV epidemic on this measure of aggregate fertility was a very small decrease, from 0.235 to 0.226. We also simulated the expected total fertility for a woman from age 15 (when she is assumed not to have given birth) until her 40th birthday,

with no HIV and a 15% prevalence rate. Again, the result is a very small decrease in fertility, from 5.40 children to 5.31 children.[20]

6 Robustness analysis

How can we know that the effects of HIV on fertility are due to changes in reproductive behaviour and not just to biological differences in fertility between HIV-positive and HIV-negative women? For a sub-sample of the respondents in MDHS 2004, there is information about HIV status. We re-estimate the fourth specification in Table 4, while controlling for HIV status (Table 6, column 1) and using only the sub-sample of HIV-negative women (Table 6, column 2). Any remaining effects of HIV rates in these specifications must indicate that women, regardless of their HIV status change their reproductive behaviour. Of course, we do not know when the HIV-positive women got infected. The HIV-status variable, therefore, captures the difference in fertility between women who were HIV negative in 2004 and those who were HIV positive in 2004, and its coefficient is, therefore, hard to interpret in a meaningful way. Similarly, the sample of HIV-negative women excludes women who were HIV-positive in 2004 but might have been HIV negative in earlier years.

Consistent with the findings in various other studies, women who were HIV-positive in year 2004 are less likely to give birth in any given year than HIV-negative women (specification 1). The odds ratio of giving birth for women who were HIV positive in 2004 compared to HIV-negative women is 0.81. So, as in Juhn et al. (2008), HIV-infected women on average had about 20% lower probability to give birth. For the district HIV prevalence interaction terms, the general pattern is the same as in Table 4. Young women with no prior births are more likely to give birth where HIV prevalence is higher, whereas the oldest women are less likely to give birth to their first child. Women who already have children generally have a lower probability of giving birth where the district HIV prevalence is higher. The differences in estimated coefficients in these two specifications, compared to specification 4 in Table 4, appear to be due to restricting the sample to women with HIV-status information rather than to controlling for HIV status (column 1) or excluding HIV-positive women (column 2). Estimation of specification 4 in Table 4 only on the sub-sample of women with HIV-status information, i.e. without controlling for HIV status, gives results almost identical to those in specification 1 in Table 6 (available from the authors upon request).

Table 6 also includes a specification where the relative household-wealth variables have been excluded (specification 3), because, strictly speaking, we only know the relative wealth of the woman's household in the survey year. The specification 4 also excludes education variables as education could be

[20]The TFR is usually calculated for women 15–49, which of course gives a larger number.

Table 6 Logit estimates of the fertility effect of the HIV epidemic: robustness analysis

	(1) HIV-status control	(2) HIV-negative only	(3) No hh wealth controls	(4) No hh wealth or education	(5) Children alive (not born)	(6) Random effects estimator
Wealth quintile dummies	Yes	Yes	No	No	Yes	Yes
Primary and secondary school dummies	Yes	Yes	Yes	No	Yes	Yes
HIV-positive in 2004	−0.193[c]					
	(0.054)					
District HIV interaction terms						
No prior births age 15–19[a]	0.010[a]	0.009	0.002	0.003	0.002	0.002
	(0.005)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
No prior births age 20–24[a]	0.026[c]	0.032[c]	0.020[c]	0.017[c]	0.015[c]	0.020[c]
	(0.008)	(0.009)	(0.003)	(0.003)	(0.002)	(0.003)
No prior births age 25–29[a]	−0.023	−0.022	−0.005	−0.006	−0.007[a]	−0.005
	(0.016)	(0.020)	(0.005)	(0.004)	(0.004)	(0.005)
No prior births age 30–34[a]	−0.046[a]	−0.061[a]	−0.043[c]	−0.043[c]	−0.043[c]	−0.043[c]
	(0.028)	(0.033)	(0.011)	(0.011)	(0.008)	(0.011)
No prior births age 35–39[a]			−0.107[c]	−0.105[c]	−0.089[c]	−0.107[c]
			(0.033)	(0.033)	(0.019)	(0.033)
One or two prior births age 15–19[a]	0.003	0.006	0.004[a]	0.004[a]	0.005[a]	0.004[a]
	(0.007)	(0.007)	(0.002)	(0.002)	(0.003)	(0.002)
One or two prior births age 20–24[a]	−0.001	0.003	−0.004[a]	−0.004[a]	−0.005[b]	−0.004[b]
	(0.005)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
One or two prior births age 25–29[a]	0.006	0.008	−0.004	−0.004[a]	−0.000	−0.004
	(0.007)	(0.007)	(0.002)	(0.002)	(0.002)	(0.002)
One or two prior births age 30–34[a]	−0.038[b]	−0.051[c]	−0.036[c]	−0.036[c]	−0.022[c]	−0.036[c]
	(0.016)	(0.019)	(0.004)	(0.004)	(0.003)	(0.004)
One or two prior births age 35–39[a]	−0.093[b]	−0.124[c]	−0.071[c]	−0.070[c]	−0.046[c]	−0.071[c]
	(0.038)	(0.048)	(0.009)	(0.009)	(0.005)	(0.009)
Three or four prior births age 15–19[a]	0.039[b]	0.039[b]	0.007	0.007	0.007	0.007
	(0.015)	(0.017)	(0.009)	(0.009)	(0.018)	(0.009)
Three or four prior births age 20–24[a]	−0.019[b]	−0.016[a]	−0.016[c]	−0.015[c]	−0.020[c]	−0.016[c]
	(0.008)	(0.009)	(0.003)	(0.003)	(0.004)	(0.003)
Three or four prior births age 25–29[a]	−0.007	−0.003	−0.001	−0.000	−0.005[b]	−0.001
	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
Three or four prior births age 30–34[a]	0.002	0.003	0.000	0.000	0.005[b]	0.000
	(0.009)	(0.010)	(0.003)	(0.003)	(0.003)	(0.003)
Three or four prior births age 35–39[a]	−0.038[c]	−0.053[c]	−0.022[c]	−0.022[c]	−0.008[b]	−0.022[c]
	(0.014)	(0.015)	(0.004)	(0.004)	(0.004)	(0.004)
Five or more prior births age 15–19[a]			−0.830[c]	−0.825[c]	−0.731[c]	−0.854[c]
			(0.230)	(0.230)	(0.242)	(0.231)
Five or more prior births age 20–24[a]	0.018	0.037[a]	−0.019	−0.017	−0.029	−0.019[a]
	(0.021)	(0.023)	(0.013)	(0.013)	(0.024)	(0.011)
Five or more prior births age 25–29[a]	0.001	0.001	−0.012[c]	−0.011[c]	−0.018[c]	−0.012[c]
	(0.009)	(0.011)	(0.003)	(0.003)	(0.005)	(0.003)

endogenous if girls make joint decisions on fertility and education early in life. Excluding the wealth and schooling variables has no substantial effect on estimated coefficients of other variables.

Table 6 (continued)

	(1) HIV-status control	(2) HIV-negative only	(3) No hh wealth controls	(4) No hh wealth or education	(5) Children alive (not born)	(6) Random effects estimator
Five or more prior births age 30–34[a]	0.001 (0.007)	0.001 (0.007)	−0.000 (0.002)	0.000 (0.002)	−0.005 (0.003)	−0.000 (0.002)
Five or more prior births age 35–39[a]	−0.020[a] (0.011)	−0.022[b] (0.011)	−0.005 (0.003)	−0.004 (0.003)	−0.006[a] (0.003)	−0.005 (0.003)
Variance of unobserved individual effects						0.000 (0.000)
Observations	17,185	14,233	148,085	148,085	148,020	148,020
Log pseudo-likelihood	−8,825.20	−7,369.41	−7,4505.5	−7,4590.6	−7,4737.9	−7,4465.5

Dependent variable is birth/no-birth; all estimations include a constant, age group dummies, prior-birth dummies, an urban-residence dummy, the number of siblings, child mortality among siblings, ethnicity dummies, year dummies and district dummies. Robust standard errors, clustered at the level of the sampling cluster, are in parentheses. In (5), dummies for Number of surviving children are used instead of dummies for number of children born, both separately and when interacted with district HIV prevalence and age dummies. Random effects are at the level of the individual *hh* household

[a]Significant at 10%

[b]Significant at 5%

[c]Significant at 1%

It could also be that it is not the number of children the woman has already given birth to that matter for her will to give birth, but the number of surviving children. Specification 5, therefore, uses information about the number of surviving children, rather than the number the woman has earlier given birth to (both for dummies and for HIV prevalence interaction terms). Again, this change has no substantial effect on the estimated coefficients.

Finally, we allowed for unobserved individual heterogeneity by using the random effects logit estimator, reported as specification 6 in Table 6. The variance of unobserved individual effects is not significantly different from zero, and again, the change of model has no substantial impact on the estimated coefficients. Thus, the women effect is similar for all, or at least the great majority, of the women, implying that insofar as there are systematic differences across women in the propensity to give birth, our explanatory variables capture this. The inclusion of woman unobserved effects is, therefore, unimportant.

7 Summary and concluding remark

We evaluated the impact of HIV/AIDS on reproductive behaviour in the general female population in Malawi, i.e. HIV-negative as well as HIV-positive women. In contrast to earlier studies on changes in reproductive behaviour due to HIV/AIDS, we allowed for heterogeneous responses depending on the woman's age and number of prior births.

Using retrospective birth information in the 2000 and 2004 Malawi Demographic and Health Surveys, we constructed a panel of yearly observations from 1980 to the survey year for each woman. The birth history was then modelled as a discrete time process, allowing for dependence on recent district HIV prevalence rates, the woman's earlier birth history and other individual and communal characteristics. There is a possibility that sexual behaviour affects both the spread of the HIV epidemic and child-bearing, so we controlled for unobserved heterogeneity across districts and ethnic groups. Hence, we conclude that the results are not due to differences in sexual and reproductive norms or behaviour across districts and ethnic groups. Nor are they due to changes in reproductive behaviour over time, as we included year dummies. Furthermore, we did random effects estimation to ensure that the results are not confounded by unobserved individual heterogeneity. To make sure that our results are due to behavioural changes in the general female population and not due to biological differences in fertility between HIV-positive and HIV-negative women, we controlled for the HIV status in 2004 of a subsample of women, and we re-estimated our model only on women who were HIV negative in 2004.

The probability that a young woman would give birth to her first child increases with the district HIV prevalence rate, whereas the probability that older women would give birth decreases. An increase in district HIV from 0% to 15% would raise the probability of a first birth for women aged 20–24 by five percentage points, whereas the probability of another birth would fall by almost five percentage points for those who have at least three children. For women over 29 years, the probability of giving birth to a first, second, or third child, would decrease by seven to 15 percentage points, and the probability of women aged 35–39 giving birth to her fourth or fifth child would decrease by about five percentage points. This suggests that young women may seek to give birth earlier when the probability of being HIV positive is lower or that women give birth earlier as a consequence of marrying and establishing supposedly monogamous relationships at a younger age, as both women and men attempt to reduce the risk of becoming infected.

The changes in age-specific fertility have a small negative impact on actual aggregate fertility, as measured by the probability that a representative woman among the 15–39-year-olds should give birth within a year, which was estimated to decrease by less than one percentage point. Similarly, the effect on women's expected total number of children appears to be tiny: the total fertility rate for women aged 15 to 40 is reduced by 0.09 children.

Even though the effects on a woman's total number of children may be very small, the HIV epidemic is likely to change the timing of births, which might affect population growth, demand for schooling and child welfare. All else equal, the change in the distribution of fertility across age groups should have a positive impact on net fertility, which is what matters for population growth and the evolution of the dependency ratio. However, as there is a negative impact from increased mortality due to HIV/AIDS, the total effect on net fertility is uncertain. The increase in the number of young mothers

is perhaps more worrying, since it is likely to impact negatively on female education beyond primary school. It is harder to predict the effects on child welfare. The younger the women, the smaller the risk that she is HIV positive, so mother-to-child transmission of HIV should decrease for a given number of births per women. Similarly, the risk that the women should die in AIDS while her off-springs are very young should be smaller, reducing the expected number of AIDS orphans. And if the HIV epidemic also leads to fewer births per women, this further decreases mother-to-child HIV transmission and the number of AIDS orphans. On the other hand, the change in fertility might raise child mortality if women start child-bearing before the age of 20, since early motherhood increases the risk of neonatal and infant mortality (NSO and ORC Macro 2005).

Our results for the overall impact of HIV/AIDS on fertility thus differ from Kalemli-Ozcan (2009) and Young (2007). We do not find that HIV/AIDS increases fertility, as Kalemli-Ozcan (2009) does using macro data. And our results do not support the findings of Young (2007) of a strong negative effect on fertility in Sub-Saharan countries. Our results are, however, consistent with Juhn et al. (2008). They report that HIV/AIDS has no or a small effect on aggregate fertility but that it reduces fertility among HIV-positive women by 20% for direct physiological reasons. We estimate that women who were HIV-positive in 2004 had 20% lower probability of giving birth. However, in contrast to Juhn et al. (2008), we find a behavioural response in the general female population when conditioning the fertility response on the woman's age and number of prior births. And it is the heterogeneity in the fertility response that explains the lack of impact on aggregate fertility.

Acknowledgements We would like to thank Lennart Flood, Måns Söderbom, Rick Wicks and two anonymous referees for valuable comments. We also thank Sida/Sarec for financial support.

References

Angeles G, Guilkey DK, Mroz TA (2005) The determinants of fertility in Rural Peru: program effects in the early years of the national family planning program. J Popul Econ 18(2):367–389

Barmby T, Cigno A (1990) A sequential probability model of fertility patterns. J Popul Econ 3(1):31–51

Becker G, Lewis HG (1973) On the interaction between the quantity and quality of children. J Polit Econ 81(2):S279–S288

Bell C, Devarajan S, Gersbach H (2004) Thinking about the long-run economic costs of AIDS. In: Haacker M (ed) The macroeconomics of HIV/AIDS. International Monetary Fund, Washington

Bloom D, Mahal A (1997) Does the AIDS epidemic threaten economic growth. J Econom 77(1):105–124

Chimbwete C, Watkins SC, Msiyaphazi Zulu E (2005) The evolution of population policies in Kenya and Malawi. Popul Res Policy Rev 24(1):85–106

Corrigan P, Glomm G, Mendez F (2005) AIDS crisis and growth. J Dev Econ 77(1):107–124

Doctor HV, Weinreb AA (2003) Estimation of AIDS adult mortality by verbal autopsy in Rural Malawi. AIDS 17(17):2509–2513

Durevall D, Lindskog A (2009) How does communal HIV/AIDS affect fertility?: evidence from Malawi. Scandinavian Working Papers No 369, University of Gothenburg

Fabiani M, Nattabi B, Ayella E, Ogwang M, Declich S (2006) Differences in fertility by HIV serostatus and adjusted HIV prevalence data from an antenatal clinic in Northern Uganda. Trop Med Int Health 11(2):182–187
Fink G, Linnemayr S (2008) HIV, education, and fertility: long-term evidence from Sub-Saharan Africa. Harvard School of Public Health, Cambridge, Mimeo
Fortson J (2009) HIV/AIDS and fertility. Am Econ J Appl Econ 1(3):170–194
GoM (2007) Malawi HIV and AIDS monitoring and evaluation report 2007. Government of Malawi, Lilongwe
Gray RH, Wawer MJ, Serwadda D, Sewankambo N, Li C, Wabwire-Mangen F, Paxton L, Kiwanuka N, Kigozi G, Konde-Lule J, Quinn TC, Gaydos CA, McNairn D (1998) Population-based study of fertility in women with HIV-1 infection in Uganda. Lancet 351(9096):98–103
Grieser M, Gittelsohn J, Shankar AV, Koppenhaver T (2001) Reproductive decision making and the HIV/AIDS epidemic in Zimbabwe. J South Afr Stud 27(2):225–243
Juhn C, Kalemli-Ozcan S, Turan B (2008) HIV and fertility in Africa: first evidence form population based surveys. NBER Working Paper No 14248
Kalemli-Ozcan S (2003) A stochastic model of mortality, fertility and human capital investment. J Dev Econ 70(1):103–118
Kalemli-Ozcan S (2009) AIDS, reversal of the demographic transition and economic development: evidence from Africa. NBER Working Paper No 12181
Lewis JC, Ronsmans C, Ezeh A, Gregson S (2004) The population impact of HIV on fertility in Sub-Saharan Africa. AIDS 18(suppl 2):35–43
Lorentzen P, McMillan J, Wacziarg R (2008) Death and development. J Econ Growth 13(2):81–124
Magadi MA, Agwanda A (2007) The link between HIV/AIDS and recent fertility patterns in Kenya. Measure Evaluation Working Paper No 07–92, Caroline Population Centre, University of North Carolina
McDonald S, Roberts J (2006) AIDS and economic growth: a human capital approach. J Dev Econ 8(1):228–250
Measure DHS (2008) Statcompiler database. http://www.statcompiler.com
Morah E (2007) Are people aware of their HIV-positive status responsible for driving the epidemic in Sub-Saharan Africa: the case of Malawi. Dev Policy Rev 25(2):215–242
Neyman J, Scott E (1948) Consistent estimates based on partially consistent observations. Econometrica 16(1):1–32
Noël-Miller CM (2003) Concerns regarding the HIV/AIDS epidemic and individual childbearing: evidence from Rural Malawi. Demogr Res 1(10):319–348
NSO (National Statistical Office of Malawi) (1993) Malawi population and housing census 1987, vol II. NSO, Zomba
NSO (2002) Census analytical report. NSO, Zomba
NSO and ORC Macro (2001) Malawi demographic and health survey 2000—final report. NSO, Zomba
NSO and ORC Macro (2005) Malawi demographic and health survey 2004—final report. NSO, Calverton
NSO and UNICEF (2008) Multiple indicator survey 2006: Malawi. NSO, Zomba
Ntozi J (2002) Impact of HIVAIDS on fertility in Sub-Saharan Africa. Paper prepared for the fourth meeting of the follow-up committee on the implementation of the DND and the ICPD-PA Yaounde, Economic Commission for Africa in collaboration with UNFPA
Oladapo OT, Daniel OJ, Odusoga OL, Ayoola-Sotubo O (2005) Fertility desires and intentions of HIV-positive patients at a Suburban specialist center. J Natl Med Assoc 97(12):1672–1681
Papageoriou C, Stoytcheva P (2008) What do we know about the impact of AIDS on cross-country income so far? Working Paper No 2005-01 Department of Economics, Louisiana State University
Rutenberg N, Biddlecom AE, Kaona FAD (2000) Reproductive decision-making in the context of HIV and AIDS: a qualitative study in Ndola, Zambia. Int Fam Plann Perspect 26(3):124–130
Rutstein SO, Johnson K (2004) The DHS wealth index. DHS Comparative Report No 6 ORC Macro. NSO, Calverton
Santaeulalia-Llopis R (2008) Aggregate effects of AIDS on development. Washington University in Saint Louis, Saint Louis, Mimeo

Soares RR (2005) Mortality reductions, educational attainment, and fertility choice. Am Econ Rev 95(3):580–601
Soares RR (2006) The effect of longevity on schooling and fertility: evidence from the Brazilian demographic and health survey. J Popul Econ 19(1):71–97
Terceira N, Gregson S, Zaba B, Mason P (2003) The contribution of HIV to fertility decline in rural Zimbabwe, 1985–2000. Popul Stud 57(2):149–164
Ueyama M, Yamauchi F (2009) Marriage behavior response to prime-age adult mortality—evidence from Malawi. Demography 46(1):43–63
Werker E, Ahuja A, Wendell B (2006) Male circumcision and AIDS: the macroeconomic impact of a health crisis. Harvard Business School Working Paper No 07-025
Westoff CF, Cross AR (2006) The stall of the fertility transition in Kenya. DHS Analytical Studies No 9. ORC Macro, Calverton
Yeatman SE (2009) The impact of HIV status and perceived status on fertility desires in Rural Malawi. AIDS Behav 13(suppl 1):12–19
Young A (2005) The gift of dying: the tragedy of AIDS and the welfare of future African generations. Q J Econ 120(2):423–466
Young A (2007) In sorrow to bring forth children: fertility amidst the plague of HIV. J Econ Growth 12(4):283–327
Zaba B, Gregson S (1998) Measuring the impact of HIV on fertility in Africa. AIDS 12(suppl 1):41–50

Paper III

HIV/AIDS, Mortality and Fertility: Evidence from Malawi

Dick Durevall and Annika Lindskog

Department of Economics
School of Business, Economics and Law
University of Gothenburg

Abstract

This paper studies the effect of HIV/AIDS on actual and desired fertility in rural Malawi, using the 2004 Demographic and Health Survey. The focus is on HIV-negative women and men, and behavioral responses in the general population. To avoid feedback effects, lagged prime-age mortality is used as a proxy for HIV/AIDS, and to control for time-invariant factors influencing both fertility and prime-age mortality, pre-HIV district fertility is used. We find a positive behavioral fertility response to mortality increases. Moreover, actual fertility responds positively to male mortality but negatively to female mortality, while women's and men's desired fertility respond negatively to mortality. These findings are consistent with an insurance and old-age security motive for having children among rural Malawian women. When a woman risks death before her children grow up, the value of children is low, and when the risk of husband's death is high, the value of children is high. We also find that the positive fertility response is limited to younger women, with no discernable age-pattern in desired fertility effects. Possible reasons are early marriage to reduce risk of HIV-infections and having babies early to reduce the risk of giving birth to HIV-infected babies.

Keywords: AIDS, demand for children, fertility, gender, HIV, mortality, prime-age adult mortality.

JEL classification: I10, J13, O12

Acknowledgement
We would like to thank Sida/Uforsk for financial support, and Måns Söderbom and Rick Wicks for very useful comments.

1. Introduction

In the worst affected countries in Sub-Saharan Africa, HIV rates have been over 10% among adults for more than two decades, generating many-fold increases in prime-age mortality (Oster, 2010; UNAIDS, 2010). Such high HIV-infection and mortality rates are likely to affect behavior in a variety of ways. One question that has attracted attention recently is how HIV/AIDS, and the associated increases in mortality, influences fertility. Since there are many youths and they are increasing rapidly, changes in fertility have strong effects on population growth, dependency ratios, and the number of new entrants to the labor market, all of which can be decisive for future development (Young, 2005; Kalemli-Ozcan, 2010).

Several studies estimate the effect of HIV/AIDS on fertility in Sub-Saharan Africa, and there is by now ample evidence that the physiological effects of HIV reduce fertility by about 20% to 40% (Lewis et al., 2004). Although this effect is substantial, it is limited to infected women, and the resulting impact on fertility in general is marginal even in countries with high infection rates. On the other hand, the evidence on the overall response, including uninfected women, is inconclusive. For example, Kalemli-Ozcan (2010) finds a positive effect with cross-section data at the country and region levels, but inconclusive results when within-country variation is used, Boucekkine et al. (2009), using country-level data, find a substantial negative impact, while Fortson (2009), using repeated cross sections for 12 countries and regional total fertility rates, finds no impact at all. According to Kalemli-Ozcan (2010), one reason for the diverse findings is the use of bad country-level data; newly available individual data with HIV tests should be used instead. Another plausible reason is heterogeneity, i.e., responses vary across countries, regions, ethnic groups, age groups, etc. With many plausible mechanisms linking HIV/AIDS and fertility, it seems unlikely that they generate a homogenous response, i.e., that they are equally important in all countries. The cross-country average effects may therefore be uninformative about responses in specific contexts.

This paper analyzes the impact of AIDS-related prime-age mortality on actual and desired fertility in rural Malawi among HIV-negative men and women, using ordered probit models. By restricting the sample to those who are uninfected, we avoid mixing up behavioral and physiological effects. In contrast to most earlier empirical studies, we focus on one country, Malawi, and go beyond average effects, analyzing differences in response due to gender-

specific district prime-age mortality and age-specific effects.[1] The data on actual fertility, ideal number of children, i.e. desired fertility, and HIV infection, come from the 2004 Malawi Demographic and Health Survey (MDHS). Our HIV/AIDS indicator is district prime-age mortality rates, obtained from the 1998 Population Census; the survey data on district HIV prevalence is from 2004, which is after actual fertility had taken place. Ueyama and Yamauchi (2009) also use prime-age mortality as proxy for HIV/AIDS in a study on marriage behavior.

The spread of HIV has not been random, and in the ideal case we would have had an instrument for mortality, but it is difficult to find a credible one. We avoid the problem of feedback effects by using lagged mortality rates. However, it is possible that HIV spread rapidly in districts that had high fertility rates, not the least since both fertility and HIV infection are outcomes of sexual intercourse. We therefore include district fertility in 1987 in the regressions, which is before HIV/AIDS had become common, controlling for factors that affected fertility in the same way in 1987 as in 1998, i.e. time-invariant factors. This proves to be important as it changes the sign of the total fertility effect from negative to positive.[2]

Our results indicate that prime-age mortality, and thus HIV/AIDS, has a positive impact on fertility due to behavioral change; it contributed to a ten percent increase during the period analyzed, mid-1999 – mid-2004. However, when the model is estimated with the whole sample, including HIV-positive women and those who refused to be tested, the impact is close to zero and insignificant, possibly due to the combined behavioral and physiological effects. Thus, we fail to find support for the hypotheses that HIV/AIDS either sharply reduces or increases fertility.

We find two types of heterogeneity. First, female mortality has a negative effect on actual fertility, while male mortality has a positive and larger effect. Second, the youngest women tend to increase fertility due to mortality, while there is no significant response among women over 29 years. Using indicators of the HIV epidemic instead of mortality and different empirical method, Durevall and Lindskog (2011), also find a positive effect on young women's fertility, but a negative and significant effect on older women's fertility.

[1] Fink and Linnemayr (2009), analysing micro-data from five African countries, but not Malawi, find that better educated women reduce fertility as a response to the HIV epidemic, while less educated women increase it. We also tested for different responses depending on the level of education, but could not confirm the findings in Malawi. A possible explanation is that our sample only includes the rural population and that there are few people with tertiary education. The results from the regressions are available on request.

[2] The change of sign indicates that districts with lower fertility were hit harder by the epidemic.

The desire to have more children when male mortality is high, and thus where it is more likely that husbands are HIV positive, could be an outcome of women's concern with insurance and old-age support. When female mortality is high, and it is more likely that women are HIV positive, women have a lower probability to survive until their children become adults, reducing the need of insurance and old-age support. Although the positive impact of male mortality could be an indication of male influence on fertility, this is contradicted by the fact that men's ideal number of children is reduced by male mortality (just as women's ideal number of children is reduced by female mortality). The finding of increased fertility among young women is not matched by a corresponding increase in young women's ideal number of children. Thus, we interpret it as earlier birth-giving: HIV/AIDS induces young women (or possibly men) to marry earlier and give birth to their first child sooner because they are less likely to be become HIV positive and to infect their newborns.

The following section outlines various mechanisms through which HIV/AIDS might impact on fertility and briefly reviews earlier studies on Sub-Saharan Africa. Section 3 describes the recent development of HIV/AIDS and fertility in Malawi. Section 4 presents the empirical model, and Section 5 describes the data. Section 6 reports the results from the empirical analysis, and Section 7 summarizes and draws conclusions.

2. Theory and Evidence on HIV/AIDS and Fertility

In this section we first outline how HIV/AIDS might affect actual and desired fertility, and then briefly review recent studies on HIV/AIDS and fertility in Sub-Saharan Africa.

2.1 How does HIV/AIDS affect fertility?

HIV/AIDS affects fertility in numerous ways with no single theory embracing them all. As a heuristic device, the effects can be collected into three groups: direct physiological effects; changes in behavior to reduce the risk of HIV infection; and changes in desired number of children (i.e., preferences) for both HIV-positive and HIV-negative women and men.

The physiological channel works through several mechanisms. Among the most important ones are: higher rates of miscarriage and stillbirth; co-infection with other sexually transmitted diseases; weight loss leading to amenorrhea; and reduced frequency of intercourse because of illness and premature death of regular partner. They all point towards

reduced fertility among HIV-infected women, which is 20% to 40% less than for uninfected women (Lewis et al., 2004; Juhn et al., 2009).

The risk of HIV infection increases the expected cost of sexual contact, particularly of risky sexual behavior. Thus we should expect to see less risky sexual behavior, i.e. increased abstinence, delayed age at sexual debut, increased condom use, fewer concurrent partners, and less extra marital sex, which could translate into lower fertility. There is an ongoing debate about the nature and extent of changes in sexual behavioral induced by the HIV epidemic, but there appears to have been little change in most African societies, particularly when considering the severity of the epidemic (Glick and Sahn, 2008).[3]

A response that would reduce the risk of infection is to marry and establish a (hopefully) monogamous relationship early. Ueyama and Yamauchi (2009) find that Malawian women marry earlier when prime-age mortality is higher. Women might marry young voluntarily, but it could also be that men have more bargaining power on the marriage market and decide to marry younger wives, less likely to be HIV positive. Another possible explanation is pressure on orphans to leave foster families, In Malawi, roughly 18% of the children, or 1.1 million, are considered to be orphans (UNGASS, 2010). In any case, the fertility effect is that women start childbearing earlier.

Potential parents might also wish to avoid giving birth to HIV-infected babies, who would die at an early age without anti-retroviral therapy (Young, 2005; 2007). It is thus possible that some women, instead of abstaining from having children, have them earlier, when they have a smaller probability of being HIV positive. In a qualitative study, Zimbabwean women mention the possibility of both decreasing the number of births and giving birth earlier as responses to the risk of giving birth to HIV-infected babies (Grieser et al., 2001).

There is a sizeable literature on child mortality and fertility mostly pointing towards a positive relationship (Schultz, 1997). AIDS raises child mortality, but in Malawi under-five mortality has dropped sharply in the midst of the epidemic, from 218 per 1,000 births in 1990 to 110 in 2009 (UNICEF, 2010).[4] Hence, AIDS-induced child mortality does not seem to be

[3] Oster (2005) argues that the difference in response between the homosexual Americans and heterosexual Africans could be explained by the shorter life expectancy and lower incomes in Africa, which reduce the value of staying uninfected.

[4] Child mortality declined because other factors dominated over AIDS, including increased immunization, vitamin A supplementation, and exclusive breastfeeding, and elimination of neonatal tetanus (NSO and UNICEF, 2008).

large enough to have a substantial direct effect on fertility. This does not preclude it from affecting fertility through the risk of giving birth to HIV-infected babies.

One of the most stunning effects of HIV/AIDS is the sharp increase in prime-age mortality: in Malawi it rose fourfold from the 1980s to 2000 according Doctor and Weinreb (2003). According to Soares (2005; 2006), adult mortality increases fertility through two mechanisms: First, it reduces returns to education, increasing the relative attractiveness of childbearing, where the two are seen as alternatives. Though Soares focuses on the total number of children, women would also start childbearing earlier if they get married and have children instead of continuing school. Second, parents care about the continued survival of their lineage, or at least evolution implies that they behave as if they do. This means they have more children when children's life expectancy as adults is lower. In the context of HIV/AIDS, the increase in prime-age mortality should thus raise fertility (Lorentzen et al., 2008; Juhn, et al., 2009; Kalemli-Ozcan, 2010).

Young (2005, 2007) argues that there is an indirect mechanism working in the opposite direction: HIV leads to lower fertility because mortality-induced wage increases and new job openings raise the opportunity cost of having children. Boucekkine et al. (2009) incorporate this effect in a general equilibrium model of the HIV epidemic that also accounts for adult and child mortality. The authors conclude that there is an ambiguous impact of adult mortality when a negative income effect is allowed for. Hence, the total impact on fertility is also ambiguous.

A potential effect, not included in the models, is the insurance and old-age security motive, which is likely to be a major reason for having children in Sub-Saharan Africa, particularly in rural areas (Nugent, 1985; Pörtner, 2001; Boldrin et al., 2005). Increased death rates of grown-up children thus means that more children are needed to ascertain a given number of survivors. However, HIV/AIDS will also affect parents own mortality risk. When parents expect to die before their children reach adulthood, they have less need of the insurance and old-age security, which reduces the marginal benefit of children. Ainsworth et al. (1998) briefly mention this mechanism in relation to HIV/AIDS, but to our knowledge it has not been further explored in the recent literature. The effect might be strong, since there are indications that the HIV epidemic has raised the subjective own-mortality risk in Malawi much more than the objective risk (Delavande and Kohler, 2009).

Further, the insurance and old-age security motive to have children is likely to be more important for women than men (Nugent, 1985). A typical woman is younger than her husband and lives longer, so she outlives him by several years and should expect long periods of widowhood. Moreover, job opportunities are much more limited for women than men, and in rural Malawi they often lack the knowledge to continue with commercial farming when the man dies. Another difference is that women often risk losing their property when they become widows because of land grabbing (Arrehag et al., 2006). Thus, female adult mortality, related to women's own mortality risk should reduce women's demand for children. Male adult mortality is instead related to her spouse's risk of death and the expected length of widowhood, increasing the demand for children.

2.2 Empirical studies on HIV and fertility

Until recently most studies focused on the fertility response among HIV-positive individuals or on small samples in a few cities (Setel, 1995; Notsi, 2002; United Nations, 2002). As mentioned, the main finding was that HIV infection reduces fertility due to physiological reasons. The impact on total fertility would be small, however. For example, if 15% of the women are HIV positive Malawi, as indicated by UNGASS (2010), fertility would drop by about 3% to 5%. The focus has thus shifted to the response of women in general and the overall impact of HIV/AIDS on fertility.

So far, the findings of the overall impact of HIV/AIDS are inconclusive. Using macro-panel data over the period 1960-2000, Lorentzen et al. (2008) find that adult mortality is positively associated with high fertility, and conclude that HIV/AIDS increases fertility since it increases adult mortality. Kalemli-Ozcan (2010) tests the association between HIV/AIDS and fertility in Sub-Saharan Africa using country and country-regional data and individual data. With cross-country data, she also finds that HIV/AIDS increases fertility. However, estimates with regional data and fixed effects show a negative or no significant impact of HIV/AIDS, depending on the specification.

Some studies find that HIV/AIDS has a weak or no effect on fertility. Ahuja et al. (2006) use macro data and circumcision as an instrument to identify the causal impact of HIV/AIDS. Most of the coefficients are negative, but insignificant. Fortson (2009) constructs regional total fertility trends in 12 countries and test for an effect with a difference-in-difference approach, and finds no effect. Magadi and Agwanda (2010) study the effect of communal HIV prevalence on fertility in Kenya using DHS data from 2003 with HIV tests. They fail to

find that communal HIV affects actual or desired fertility, but indictors of HIV/AIDS awareness seem to reduce overall fertility. Finally, Durevall and Lindskog (2011) construct a panel with recall data on Malawian women's birth histories, collected by MDHS 2004.They find that HIV/AIDS has a small negative effect on fertility.

Young (2007) uses DHS data from a large sample of Sub-Saharan countries, collected before HIV testing was widely available, and national HIV rates based on data on women visiting antenatal clinics. In contrast to other studies, he finds a strong negative effect; a country with a HIV-prevalence rate at 10% reduces fertility by roughly the same magnitude as moving from no education to secondary education. Young (2005) obtains similar results for South Africa.

Juhn et al. (2009) and Kalemli-Ozcan and Turan (2010) criticize Young's studies, arguing that they are flawed. By using recently available DHS data with HIV-testing, Juhn et al. (2009) are able to analyze behavioral responses among non-infected women at the country-regional level in 13 Sub-Saharan African countries. They find no significant effect. HIV/AIDS reduces fertility due to actual infections, but the overall impact is very small. Kalemli-Ozcan and Turan (2010) show that Young's (2005) finding of a large fertility decline in South Africa is due to the time period analyzed. When focusing on the period 1990-1998, for which there is information on HIV rates, fertility seems to increase.

Nonetheless, the study by Boucekkine et al. (2009), who estimate dynamic models for 39 Sub-Saharan African countries over 1980-2004, find that adult mortality reduces fertility, while child mortality increases it. They also find that only adult mortality affects the number of surviving children and conclude that overall the HIV/AIDS epidemic has unambiguous negative effect on net fertility, i.e., the number of surviving children.

The research reviewed has mainly concentrated on the effect on overall fertility and differences between infected and non-infected women. Hence, the failure to find clear-cut behavior responses might be due heterogeneity. There are a few studies that shed light on heterogeneity in age and schooling among women. Durevall and Lindskog (2011) directly test for age-specific responses in Malawi, and find that young woman give birth to their first child sooner, while older women, who have already started child-bearing, decrease their fertility. Ueyama and Yamauchi (2009), using MDHS 2004 data, find that Malawian women marry earlier if prime-age adult mortality is high. And Noël-Miller (2003), using Malawian

panel data, finds that the association between the degree of worry regarding HIV/AIDS infection and the number of births among young women is positive and negative among older women. One study focuses on schooling, Fink and Linnemayr (2009). They use data from five African countries, but not Malawi, and find that better educated women reduce fertility as a response to the HIV epidemic, while less educated women instead increase it. Hence, all these studies indicate there might be substantial, but heterogeneous, behavioral changes, to HIV/AIDS, which might counteract each other.

3. Fertility and HIV/AIDS in Malawi

In the early 1960s, the total fertility rate in Malawi was similar to those in other African and other less-developed countries. But while fertility in most other countries fell during 1960-1980, it grew in Malawi, probably because of the ideology and policy of the Malawian government under President Banda: birth control was seen as incompatible with Malawian culture (Chimbwete et al., 2005). Fertility started to fall in the mid 1980s at a rate similar to that in many other African countries: the total fertility rate declined from 7.6 in 1984, to 7.4 in 1987, and 6.5 in 1998. Nonetheless, it is still high compared to most countries, and there are indications that it has stopped falling. It has varied between 6.0 and 6.3 since 2000, implying that, on average, women in Malawi gave birth to one more child than women in the rest of Africa (NSO, 2010a).

Malawi's first AIDS case was diagnosed in 1985, when the national HIV prevalence rate should still have been very low. From then on the epidemic spread rapidly, first in the major cities, and then in rural areas. In the cities the HIV rate peaked in 1995 at 26% among women attending antenatal clinics, and then started to decline slowly. In the rural areas the rate was estimated to be 11.8% in 1999, and 10.8% in 2004 (National Aids Commission, 2004; NSO and OCR Macro, 2005). According to the most recent data, the national rate was 11% in 2009, during which 50,000 people are estimated to have died from AIDS, out of a population of 13 million (UNAIDS, 2010).

HIV/AIDS has thus been prevalent in Malawi for over 25 years, and the epidemic has increased prime-age mortality about four times, i.e., three out of four deaths among prime-age adults are due to AIDS (Doctor and Weinreb, 2003). As a result, knowledge about AIDS is widespread. In fact, already in the MDHS carried out in 1992, about 90% of respondents

had heard about the disease, rising to 99% in the 2000 MDHS (NSO and Macro International, 1994; NSO and ORC Macro, 2001). So if HIV/AIDS affects decision-making about childbearing, this should be measurable.

One of the striking features of the epidemic is its differential impact on men and women: In 2004, close to 60% of the infected adults were women. Furthermore, male and female HIV rates vary widely across districts. For example, in Blantyre (with the most important commercial city), men and women have an equal probability of being HIV positive, while in Zomba (with an important university city), women have twice the probability of being affected as men (NSO and ORC Macro, 2005).

4. Empirical Model of Fertility and Inference

When analyzing the effect of the HIV epidemic on fertility, we are ultimately interested in the effect on women's complete fertility, the total number of children given birth, and possibly the timing of those births. However, since the epidemic started in earnest in the mid-1980s, it is too early to study its effects on complete birth histories. Thus, using the approach of Soares (2006), we focus on fertility during the period 1999-2004. Soares studies childbearing up to the date of the survey, treated as a function of the woman's individual choice, factors not under her control, and her age. Since we are studying fertility during a limited period, we also use prior births to control for the stage of the reproductive life cycle the woman is in. Furthermore, fertility is allowed to depend on recent information on the HIV epidemic. There is also uncertainty in the model, which captures the fact that women cannot control their fertility perfectly, for biological reasons, such as fecundity, and for social reasons, such as their partner's attitudes.

We assume that the number of births during the study period is a function of B, a latent continuous variable that indicates the propensity to have a certain number of births, where $B = N + \varepsilon$, with $N = N(n(X), t, pb)$, and ε a random term representing uncertainty. Behavior is determined by desired lifetime fertility, $n = n(X)$, where X includes individual and communal factors; by the age of the woman, t; and the number of prior births, pb. The actual number of births, N, during a given period for a woman at a certain age is

0 if $B \leq c_0$;

k if $c_{k-1} < B \leq c_k$, $k = 1-3$;

4 if $c_3 < B$;

where $c_0 - c_3$ are cut-off values and 4 is the maximum number of births observed during the period. We assume that ε is normally distributed and estimate this as an ordered probit model. The probability that a women will not give birth to any children during the period is then

$$\begin{aligned} P(0) &= P(B \leq c_0) \\ &= P(\varepsilon \leq c_0 - N) \\ &= \Phi(c_0 - N) \end{aligned} \tag{1}$$

where $\Phi(.)$ indicates the standard normal distribution function. The other probabilities can be specified as

$$\begin{aligned} P(1) &= \Phi(c_1 - N) - \Phi(c_0 - N) \\ P(2) &= \Phi(c_2 - N) - \Phi(c_1 - N) \\ P(3) &= \Phi(c_3 - N) - \Phi(c_2 - N) \\ P(4) &= 1 - \Phi(c_3 - N) \end{aligned} \tag{2}$$

The values of $c_0 - c_3$ are estimated as parameters in the model, together with the coefficients in $N = N(n(X), t, pb)$.

Our main explanatory variable, district prime-age mortality, only varies across Malawi's 27 districts. It is, thus, essential to account for within-district dependence in the error term (Moulton, 1990; Bertrand et al., 2004). Within-cluster dependence is usually accounted for with clustered standard errors. However, estimation of clustered standard errors, like other 'sandwich type' standard errors, relies on large-sample asymptotics, and thus requires a large number of clusters for correct inference. We use a procedure proposed by Cameron et al. (2008) for inference with few clusters, the cluster bootstrap-t procedure.[5]

The cluster bootstrap-t procedure works in the following way. In each iteration *j*, the bootstrap randomly samples, with replacement, the number of clusters in the original sample. Thereafter the regression model is run, and the Wald-statistic $w_b^* = \left(\hat{\beta}_{j,b}^* - \hat{\beta}_b\right) / s_{\hat{\beta}_{j,b}^*}$ is calculated, where $\hat{\beta}_b$ is the original sample parameter, $\hat{\beta}_{j,b}^*$ is the parameter of the *j*th

[5] They also suggest the wild-bootstrap, which is a modified residual bootstrap. In an ordered probit model it is, however, not appropriate to bootstrap 'residuals'.

iteration, and $s_{\hat{\beta}^*_{j,b}}$ is the cluster corrected standard error of the parameter in the *j*th iteration. The null-hypothesis that the parameter equals zero is then rejected at α level of significance if the original sample Wald-statistic *w* is either smaller than $W^*_{[\alpha/2]}$ or larger than $W^*_{[1-\alpha/2]}$.

5. Data and Variables

The main source of data is the nationally representative MDHS carried out in 2004 (Measure DHS, 2010). Apart from fertility-related information and data on a range of characteristics of the respondents and their households, the 2004 MDHS contains HIV status for a subsample, the first nationally representative survey of HIV prevalence.

We focus on rural areas for two reasons. First, the insurance and old-age security motive for having children, central to our analysis, is likely to be much stronger in rural than urban areas (Nugent, 1985). Second, HIV was virtually non-existent in rural areas in the mid-1980s (UNGASS, 2010) and thus should not have affected district fertility in 1987, an important control variable to deal with possible endogeneity. In addition, about 85% of the population in Malawi lives in the rural areas (NSO, 2010b).

Furthermore, we use the sample of knowingly HIV-negative women, i.e. women from whom a negative sero-prevalence blood sample was successfully collected. This is to avoid mixing up behavioural responses in the general female population with physiological effects among HIV-positive women. Although we know that they are HIV negative, very few of them did, since it was uncommon for people to go for HIV testing before AIDS had developed (Morah, 2007). Thus, we should capture the effects of an uncertain own HIV status on fertility preferences.

Since only about 80% of women agreed to do the HIV-test, there is a potential sample-selection problem. In the final 2004 MDHS report, the issue of potential response bias is investigated by comparing observed and predicted HIV rates for different groups of people (NSO and ORC Macro, 2005). With the exception of the capital Lilongwe, observed and predicted rates differ little. Durevall and Lindskog (2009) also compare observed characteristics of respondents who provided the blood sample and those who refused, but only find minor differences. In any case, since we focus on HIV-negative women, refusal to test among infected women seems unlikely to bias our results.

Our key dependent variable is realised fertility, which we measure as the number of births during the five-year period preceding data collection; i.e. from approximately mid-1999 to mid-2004.[6] The period was chosen to start after our measurement of mortality, which is for one year prior to the census, carried out in 1998. On average women gave birth to almost one child during the period (see Table A1).

As a complement to estimations on realised fertility, and to provide information about whether changes in realised fertility are related to changes in the desired number of children, we also estimate models with desired fertility as the dependent variable. It is measured by respondents' stated ideal number of children. Women and men were asked how many children they would have liked to have if it was possible to go back in time and choose freely. As with many survey questions of a subjective nature, it is not obvious that all persons understood the question correctly, and it could be biased in favour of the number of children they already have. However, we are interested in differences across women and men, not the actual number. An important difference to realised fertility is that the ideal number of children could be understood as indicating the number of surviving children, not the number of children born. On average both women and men want to have four children (Table A1).

Our measure of mortality is district prime-age mortality rates from the 1998 census, differentiated by gender.[7] Prime-age mortality is the number of deaths per thousand individuals aged 30 to 49 years during 1997-1998 in rural areas of the district (data provided by the Malawi National Statistical Office). The mean district mortality rate is 17 deaths per 1000, varying from 7 to 36 (Table A2).

There are several advantages with the 1998 prime-age mortality rate as an indicator of HIV/AIDS compared to HIV prevalence based on DHS data, used in several other studies. First, the mortality data predates the study period, so any feed-back effects from fertility in 1999-2004 to mortality in 1998 are highly improbable.[8] Second, deaths are directly observable, as opposed to HIV status, meaning that people could act on them. Even if people are not aware of precise mortality rates and how these differ between men and women, they

[6] Tables A1 and A2 in Appendix report summary statistics for all variables used in the analysis.
[7] The choice of indicator for HIV/AIDS varies in the literature. Some examples are: national HIV-rates obtained from antenatal clinics (Young, 2005, 2007; Kalemli-Ozcan, 2010), district HIV-rates obtained from antenatal clinics (Durevall and Lindskog, 2011), AIDS deaths (Kalemli-Ozcan, 2010), death rates based on mortality of siblings reported in DHSs (Fortson, 2010), and regional HIV-rates from DHSs (Juhn et al., 2009).
[8] A serious potential drawback in studies using current HIV prevalence rates, such as Juhn et al. (2009) and Magadi and Agwanda (2010) is that women might actually have become infected while getting pregnant.

are able to observe prime-age adult deaths in their surroundings. People die due to other reasons than HIV of course, but among prime-age adults in heavily affected countries AIDS is the leading cause of death (Oster, 2010). And as noted earlier, AIDS probably caused three out of four deaths in Malawi in the age group (Doctor and Weinreb, 2003). Finally, the census district mortality data should be of reasonably good quality since it is based on many observations.

In the ideal case, we would have had an instrument for mortality; however it is difficult to find a credible one. We avoid the problem of feedback effects by the use of lagged mortality rates. However, it is possible that districts that always had high fertility rates were also harder hit by HIV, not the least since both fertility and HIV infection are outcomes of sexual intercourse. We therefore include pre-HIV/AIDS district fertility in regressions, measured by the ratio of the number of births to women in reproductive age (15-49) in each district, using data from the 1987 population census. Including 1987 district fertility in the regressions proves to be important as it changes the sign of the total mortality effect from negative to positive.[9] In this way we control for factors that were time-invariant between 1987 and 1998.

We also control for the woman's age and prior number of births. Age enters our model as seven dummy variables for the age groups 15-19, 20-24, etc., up to 45-49 (and 50-54 for men). Age at the time of the survey is used, with fertility measured during the previous five years. Prior number of births is the total up to the beginning of the five-year period.

Economic theory suggests that income and the opportunity costs of women's time should be important determinants of desired fertility. To capture these variables, household-wealth quintiles (data on income is not collected) and the woman's educational level (no schooling or incomplete primary; complete primary; complete secondary; higher education) are included. If permanent income matters more than current income for fertility, wealth should be a good proxy (Bollen et al., 2007). Unfortunately, information on wealth and education is only available from the survey year. Since less than 10% of the women had more than 8 years of schooling, most of them had reached their completed level of schooling earlier than 1999, but endogeneity is a potential problem with the youngest group. There could also be endogeneity if there were systematic changes in relative wealth over the previous five years.

[9] Districts that always had lower fertility have thus been harder hit by the epidemic than others.

However, we checked the robustness of our results by estimating models without wealth and education, and there are only marginal differences (not reported).[10]

Norms are likely to influence reproductive and sexual behaviour and thus might affect both the spread of HIV and fertility. We control for norms using dummies for ethnic and religious affiliation. It is hoped that these variables are sufficient to capture key differences in customs of the rural population.

6. Empirical Results

In this section we first test for the effect of prime-age mortality, measured by aggregated district-level mortality on overall fertility. Then we test for differences in effects due to gender-specific district mortality (sub-section 6.2) and individual age-specific fertility (sub-section 6.3).

6.1 Overall Fertility

Table 1 reports the effect of district prime-age mortality on fertility (specification 1, 4 and 5), and women's and men's desired fertility (specifications 2 and 3). The null hypothesis of a zero effect of the prime-age mortality on fertility can be rejected in favour of a positive effect, but only at the ten percent level. The magnitude of the effect of HIV/AIDS on fertility is illustrated in Table 2, which presents the predicted number of births per woman if prime-age mortality goes from its mean in 1987 (3.9 deaths per 1,000 aged 30-49) to its mean in 1998 (15.8). With the HIV epidemic, a woman in fertile age (15-49) is predicted to give birth to around 0.111 more children more during five years. However, this estimate is shaky since a 95% confidence interval would include negative values, and when the model is re-estimated with all women, i.e., the infected ones and those who were not tested, the effect of mortality is clearly insignificant (specification 4).

The positive effect on actual fertility is not matched by a corresponding increase in desired fertility. Instead, mortality is predicted to decrease women's ideal number of children by about 0.37 and men's ideal number of children by about 0.24. And these results are statistically strong; the null hypothesis of no impact of prime-age mortality is rejected at the five percent level for both women and men.

[10] The regression results are available on request.

The variable measuring pre-HIV district fertility is positively associated with women's fertility in 1999-2004, but not with women's and men's stated ideal number of children. When we do not control for pre-HIV district fertility in the regression with actual fertility, the estimated effect of mortality is negative (specification 5). Districts that had lower fertility before the HIV epidemic thus seem to have been worse affected by the epidemic than others. This is at least partially because the more densely populated Southern Region had lower fertility and more HIV in the 1990s.

Table 1: Total effect of prime-age mortality on fertility behavior – Ordered probit

	Births 1999-2004	Women's ideal number of children	Men's ideal number of children	Births 1999-2004 full sample	Births 1999-2004,
	(1)	(2)	(3)	(4)	(5)
Age 20-24	1.495***	-0.201***	-0.038	1.571***	1.503***
Age 25-29	1.559***	-0.123*	0.064	1.591***	1.560***
Age 30-34	1.225***	-0.250***	0.290***	1.213***	1.211***
Age 35-39	1.031***	-0.257***	0.396***	0.878***	1.006***
Age 40-44	0.155	-0.285**	0.293**	0.201	0.139
Age 45-49	-0.395**	-0.401***	0.242**	-0.466*	-0.414**
Age 50-54			0.177**		
Prior births	0.0368**			0.063***	0.041**
Living children		0.068***	0.133***		
Primary education	-0.0188	-0.0193	0.041	-0.026	-0.028
Secondary education	-0.359***	-0.111	-0.268***	-0.364***	-0.334**
Higher education	-1.662	-0.817	-0.467**	-0.716*	-1.728
2nd wealth quintile	-0.106**	0.0736**	-0.218***	-0.003	-0.097*
3rd wealth quintile	-0.187***	-0.0189	-0.186***	-0.045	-0.173***
4th wealth quintile	-0.212**	0.0882*	-0.373***	-0.134***	-0.194**
5th wealth quintile	-0.475***	-0.197**	-0.487***	-0.312***	-0.471***
Pre-HIV district fertility	9.863**	-4.524**	-3.798	1.407	
Prime-age mortality	13.250*	-21.130**	-17.600**	-3.611	-5.500
Observations	2143	1925	1785	10031	2149

All estimations include religion dummies, ethnicity dummies, and ordered probit cut-points (four in specifications 1, 4 and 5, and nine in specifications 2 and 3). Estimations have been done with survey weights, and standard errors have been clustered at the district.
*=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level, using a cluster bootstrap-t procedure.

Table 2: Predicted effect of the HIV/AIDS epidemic on fertility – effect of adult mortality increasing from its 1987 mean to its 1998 mean (based on specification 1 in Table 1)

	With HIV/AIDS (mortality at its 1998 mean)	Without HIV/AIDS (mortality at its 1987 mean)	Change
Number of births 1999-2004	1.0016	0.8905	0.1108
Women's ideal number of children	4.1729	4.5469	-0.3747
Men's ideal number of children	4.0456	4.2848	-0.239

The control variables have the expected effects. The coefficients of the age dummies in Table 1 (specification 1) show how age affects fertility non-linearly; it increases to 25-29, and then declines. The larger the number of births before 1999, the larger the probability of giving birth during 1999-2004. This is probably explained by differences in fertility between women that had married and started child-bearing by 1999 and unmarried women.

The effects of age on ideal number of children differ by gender; older women want fewer children than younger women, while older men want more children than younger men. Women and men with more living children also want more children, either because they have acted to fulfil their fertility intentions or because past fertility has changed (stated) fertility preferences.

As usual, education is associated with lower fertility, with a statistically significant difference between those who have less than complete primary education and those who have secondary or higher education. Household wealth is also associated with lower fertility, with statistically significant effects for those in the third, fourth and fifth quintiles. This suggests that wealthier families substitute child quality for child quantity, or that the opportunity cost of women's time is higher in wealthier households, even after controlling for their educational level.

To save space we do not report the coefficients of the ethnicity and religion dummies and cut-offs. Moreover, the coefficients of the control variables are not reported in the following estimations; their coefficients are not substantially different from those in Table 1. All results are available from the authors on request.

6.2 Effects of Gender-Differentiated Mortality

Table 3 shows coefficients from regressions that distinguish between female and male district mortality rates, while Table 4 reports predicted effects on fertility of either female mortality, male mortality, or both, due to an increase from their mean in 1987 to their mean in 1998. Differences in effects of female and male mortality can, for example, provide indirect information on the expected role of children in providing insurance and old-age security, as described earlier. If insurance and old-age security is an important motive to have children we expect women's actual and desired fertility to be positively affected by increased male mortality and negatively affected by increased female mortality.

Table 3: Effect of female- respectively male adult mortality on fertility behavior (among HIV negative people) – Ordered probit coefficients

	(1)	(2)	(3)
Female adult mortality	-8.441*	-26.43***	-7.583
Male adult mortality	17.37**	0.775	-9.676***
Tests of equality of coefficients			
Female- minus male adult mortality	-25.807**	-27.210***	2.092
Observations	2143	1925	1785

All estimations include age dummies, education dummies, household wealth dummies, religion dummies, ethnicity dummies, ordered probit cut-points (four in specification 1 and nine in specifications 2 and 3), and pre-HIV district fertility. Specification also 1 includes the number of prior births, and specifications 2 and 3 the number of living children. Estimations have been done with survey weights, and standard errors have been clustered at the district level.
*=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level, using a cluster bootstrap-t procedure.

Table 4: Predicted effects of the HIV epidemic induced increase in female and male mortality on fertility among HIV negative people

		Without HIV (mortality at its 1987 mean)	With HIV (mortality at its 1998 mean)	Change
Actual births 1999-2004	Female mortality rise	0.835	0.892	-0.056
	Male mortality rise	1.056	0.892	0.165
	Both female and male mortality rise	0.998	0.892	0.106
Women's ideal number of births	Female mortality rise	4.166	4.547	-0.381
	Male mortality rise	4.563	4.547	0.016
	Both female and male mortality rise	4.181	4.547	-0.367
Men's ideal number of children	Female mortality rise	4.199	4.285	-0.085
	Male mortality rise	4.135	4.285	-0.150
	Both female and male mortality rise	4.052	4.285	-0.234

Female and male prime-age mortality affects fertility differently. An increase in **female** mortality from its mean in 1987 to its mean in 1998 is predicted to **decrease** the average number of births 1999-2004 by 0.056 children, while a similar increase in **male** mortality is predicted to **increase** births with 0.165. The negative effect of female mortality is significant at the ten percent level, while the positive effect of male mortality is statistically significant at the five percent level. The difference between female and male mortality is statistically significant at the five percent level.

Women's ideal number of children is also affected differently by female and male mortality. If female mortality increases from its mean in 1987 to its mean in 1998, women's ideal number of children is predicted to decrease with 0.381, and the coefficient is statistically

significant at the one percent level. If instead male mortality increases from its mean in 1987 to its mean in 1998, women's ideal number of children is predicted to increase marginally, but the coefficient is not statistically different from zero. The difference between female and male mortality is statistically significant at the one percent level.

Men's ideal number of children, on the other hand, is more negatively affected by male than female mortality. An increase in male mortality from its 1987 to its 1998 mean decreases the stated ideal number of children by 0.150 children, while a similar increase in female mortality decreases it by 0.085. The coefficient for male mortality is statistically significant at the one percent level, but the one for female mortality and the difference between female and male mortality are not statistically different from zero.

6.3 Age Heterogeneity of the Fertility Response to HIV/AIDS

Tables 5 reports ordered probit coefficients from estimations of number of births and ideal number of children allowing the response to prime-age mortality to differ by age group, and Table 6 reports predicted changes in age-specific fertility due to mortality increases. There are reasons to expect that women have children earlier due to HIV/AIDS; they may marry and form hopefully monogamous relationships earlier to reduce the risk of infection; they may have children earlier to reduce the risk of giving birth to HIV-infected babies; or orphans may be pressured to leave the household early.

We have reduced the number of age groups by aggregating some of them and now use 15-19, 20-29, 30-39 and 40-49. Moreover, although fertility and women's ideal number of children respond differently to female and male mortality as reported in the previous section, we do not make this distinction. Differentiating between female and male mortality, and allowing a different response in every age group give qualitatively similar results to those reported here.[11]

[11] We also carried out estimations that allow for age-heterogeneous responses to either female mortality, male mortality, or both (available on request). The fertility response to both female and male adult mortality is more positive for younger women, and fertility still responds positively to male mortality and negatively to female mortality. The results are available on request.

Table 5 : Age heterogeneous fertility effect (on HIV-negative people) of adult mortality - Ordered probit coefficients

	(1)	(2)	(3)
Prime-age mortality *Age 15-19	36.96***	-27.55***	-35.20***
Prime-age mortality *Age 20-29	14.66**	-14.64	-11.63*
Prime-age mortality *Age 30-39	2.349	-22.69***	-23.07**
Prime-age mortality *Age 40-49	0.154	-30.00***	-2.964
Tests of equality of coefficients			
'Prime-age mortality *Age 15-19' minus 'adult mortality *Age 20-29'	22.298***	-12.906 **	-23.572***
'Prime-age mortality *Age 15-19' minus 'adult mortality *Age 30-39'	34.610***	-4.861	-12.132*
'Prime-age mortality *Age 15-19' minus 'adult mortality *Age 40-49'	36.805***	2.448	-32.241***
'Prime-age mortality *Age 20-29' minus 'adult mortality *Age 30-39'	12.313**	8.045*	11.441**
'Prime-age mortality *Age 20-29' minus 'adult mortality *Age 40-49'	13.403**	15.354**	-8.668*
'Prime-age mortality *Age 30-39' minus 'adult mortality *Age 40-49'	2.195	7.309	-20.109***
Observations	2143	1925	1785

All estimations also include age dummies, education dummies, household wealth dummies, religion dummies, ethnicity dummies, ordered probit cut-points (four in specification 1 and nine in specifications 2 and 3), and pre-HIV district fertility. Specification also 1 includes the number of prior births, and specifications 2 and 3 the number of living children. Estimations have been done with survey weights, and standard errors have been clustered at the district.

*=significant at the 10% level, **=significant at the 5% level, ***=significant at the 1% level, using a cluster bootstrap-t procedure.

Table 6: Age-specific predicted effects of the HIV epidemic on fertility (among HIV negative people)

		Without HIV (mortality at its 1987 mean)	With HIV (mortality at its 1998 mean)	Change
Actual births 1999-2004	Age 15-19	0.3736	0.2033	0.1703
	Age 20-29	1.4452	1.3175	0.1276
	Age 30-39	1.202	1.1815	0.0205
	Age 40-49	0.3949	0.394	0.0009
Women's ideal number of births	Age 15-19	4.1826	4.6732	-0.4915
	Age 20-29	4.209	4.4678	-0.2591
	Age 30-39	4.2427	4.6479	-0.4064
	Age 40-49	4.148	4.6829	-0.5342
Men's ideal number of children	Age 15-19	3.5957	4.0423	-0.4465
	Age 20-29	3.7682	3.9154	-0.1469
	Age 30-39	4.5345	4.892	-0.357
	Age 40-54	4.7758	4.8221	-0.047

The coefficients for the age groups 15-19 and 20-29 are positive, large and clearly significant, while those for older women are very small and insignificant. Young women thus give birth to considerably more children where prime-age mortality is high, while older women hardly respond. The youngest (15-19) are predicted to give birth to 0.170 more children during five years, when mortality increases from its mean in 1987 to its mean in 1998 (Table 6). Since fertility is fairly low in this group this means that fertility is almost doubled. Women in their 20s are predicted to give birth to 0.128 more children, while women in their 30s and 40s only give birth to 0.021 and 0.001 more children. The responses of the youngest and women in their 20s are statistically different from that of older women.

Nonetheless, higher mortality is associated with a desire to have fewer children irrespective of gender and age, and it is difficult to discern any age-pattern. Young people want to reduce fertility at least as much as older people. The fertility increase among the young thus seems to reflect earlier birth-giving.

7. Concluding Remarks

Understanding how HIV/AIDS affects childbearing in countries with high infection rates is of great interest, since the future course of fertility is likely to be a major determinant of population growth and economic development. This paper analyzes the impact of district prime-age mortality, a proxy for HIV/AIDS, in rural Malawi on both actual fertility and women's and men's desired fertility, among HIV-negative women and men. The aim is to shed light on the diverse findings of earlier studies by focusing on differences in fertility response due to gender-specific district mortality and age-specific effects.

We find that mortality has a positive impact on fertility among uninfected women. Those living in districts with high prime-age mortality and HIV rates give birth to more children than those living in districts with low rates. Since HIV-positive women are excluded from the sample this should be due to behavioral changes. The increase during the period analyzed, mid-1999 – mid-2004, is about 10%, but this number is uncertain since the coefficient is only significant at the ten percent level. To obtain an idea of the impact of HIV/AIDS on overall rural fertility, we can add a rough estimate of the reduction in fertility due to physiological effects, 4.5% (the decline in fecundity, 0.3, times the HIV rate among pregnant women in rural areas, 0.15). Thus fertility might have increased by about 5% during the five-year

period. Yet, if we instead use the whole sample, including HIV-positive women and those who refused to be tested, to estimate the impact, it is close to zero and insignificant.

The limited total behavioral response can be explained by another finding, substantial heterogeneity. One source of heterogeneity, which has not been studied earlier, is gender-specific adult mortality rates. Female mortality reduces fertility while male mortality increases it. Since the positive impact of male mortality is larger than the negative impact of female mortality, mortality raises fertility. However, men desire to have fewer children when male mortality is high, and women desire to have fewer children when female mortality is high. Thus, women seem to have more control over fertility than their male partner, i.e. women respond to male mortality. This appears natural too many, but contrasts with views of male dominance. Although men have more bargaining power than women in general, this is probably not the case for childbearing.[12]

Another source of heterogeneity is individual age: young women 15-29 give birth to more children where prime-age mortality is high, while the response of women over 29 is small and insignificant. Among the youngest, 15-19, fertility rose by about 80% over mid-1999 – mid-2004, albeit from a very low base, while it rose by 10% among women 20-29. Since prime-age mortality is negatively associated with women's ideal number of children for all age groups, this indicates a possible shift in timing of births, not a desire by young women to have more children.

Our findings provide some support for the prediction that fertility and mortality covary positively, made by the models of Kalemli-Ozcan (2003, 2010) and Soares (2005), and it is possible that return to education, the quality and quantity trade off, and concerns about the survival of lineage all raise fertility. It is also possible that the income effect due to improved job opportunities works in the opposite direction, reducing fertility. However, we cannot distinguish these effects. Instead we stress explanations consistent with the demonstrated patterns of heterogeneity. If insurance and old-age security is a key motive for having children, parents own increased mortality risk reduces the marginal benefit of children, but there is a difference whether mortality is more concentrated among women or men. The

[12] A woman can, for example, use contraceptives or even sterilization without asking her husbands' opinion. The use of injectable contraceptives in Malawi, which a woman can use without her husband's knowledge, increased dramatically during the 1990s, and they are by far the most popular contraceptive, especially among married women. 18% used them at the time of the survey and 41% have used them at least once (NSO and OCR Macro, 2005)

finding of a negative effect of prime-age female mortality and a positive effect of prime-age male mortality on fertility is consistent with this hypothesis; female mortality increases the risk that the woman will die, reducing the need of future support from children, while male mortality increases the risk that the husband dies early, i.e. of widowhood, and thus raises the need of support from children. This interpretation is supported by the negative impact of own-sex mortality on both women's and men's ideal number of children. Moreover, the insurance and old-age security motive to have children is likely to be of greater concern for women than men: they should expect to outlive their husbands by several years; there are fewer job opportunities for women; and some women risk losing their property when they become widows because of land grabbing (Arrehag et al, 2006).

There are several explanations for the age heterogeneity. By giving birth earlier, women reduce the risk of giving birth to HIV-infected babies, or of leaving young children orphaned. Moreover, earlier childbearing could be due to early marriage among women, as reported by Ueyama and Yamauchi (2009). This could result from both men and women aiming at establishing stable monogamous relationships to decrease the risk of HIV infection, or from comparatively older men wanting to marry young women who are less likely to be HIV positive. Early marriage and childbearing could also result from pressures on orphans to ease the burden on their foster families.

As with all observational studies, the interpretation of causality is tenuous, but the implications of our findings are straightforward. According to the 2008 population census, age specific fertility among young women has increased since the 1998 census, even though total fertility rate has declined from 6.5 to 6.0 (NSO, 2010a). The association between HIV/AIDS and fertility among young women seems to be one explanation. It is well-known that adolescent childbearing has negative health effects on both mothers and children, as well as on human capital accumulation (NSO and UNICEF, 2008). Moreover, increases in early childbearing are likely to delay the demographic transition. Hence, there is a need to focus HIV-prevention on young women. A promising approach is evaluated by Dupas (2011) who shows that information campaigns on relative risks of infection reduced teen pregnancies by 28%. Another concern is old age security for women. Although requiring further study, introducing pension for elderly has the potential to substantially reduce fertility (Boldrin, et al., 2005; Holmquist, 2010). Measures to increase women's economic independence, including land inheritance practices, might also reduce fertility.

References

Ahuja, A., B. Wendell, and E. Werker (2006), "Male Circumcision and AIDS: The Macroeconomic Impact of a Health Crisis", Harvard Business School Working Paper No. 07-025 (Revised March 2009.)

Ainsworth, M., D. Filmer, and I. Semali (1998),"The Impact of AIDS Mortality on Individual Fertility: Evidence from Tanzania" in *From Death to Birth: Mortality Decline and Reproductive Change*, Montgomery, M., and B. Cohen (eds.), National Academies Press.

Arrehag L., Durevall, D., Sjöblom M., de Vylder, S. (2006) *The impact of HIV/AIDS on Livelihoods, Poverty and the Economy of Malawi* Stockholm, Sida Studies no. 18.

Bertrand, M., Duflo, E., Mullainathan, S. (2004), "How Much Should We Trust Difference-in-Difference Estimates?", *Quarterly Journal of Economics*, 119, pp. 249-275.

Boldrin, M., De Nardi M., Jones, L.E. (2005) "Fertility and Social Security" NBER Working Paper no. 11146.

Bollen, K.A., J.L. Glanville, and G. Stecklov (2007), "Socio-economic status, permanent income, and fertility: A latent-variable approach", *Population Studies*, 61(1), 15-34.

Boucekkine, R., R. Desbordes and H. Latzer, (2009) "How do epidemics induce behavioral changes?" *Journal of Economic Growth* 14(3), 233-264.

Cameron, A.C., Gelbach, J.G., Miller, D.L. (2008), "Bootstrap-Based Improvements for Inference with Clustered Errors", *Review of Economics and Statistics*, 90, pp. 414-427.

Chimbwete, C., S.C. Watkins, and E.M. Zulu (2005), "The Evolution of Population Policies in Kenya and Malawi", *Population Research and Policy Review* 24(1), 85-106.

Delavande, A. and H.P. Kohler (200), "Subjective Mortality Expectations and HIV/AIDS in Malawi", *Demographic Research*, 20(31) 817-874.

Doctor, H.V., and A.A. Weinreb (2003), "Estimation of AIDS Adult Mortality by Verbal Autopsy in Rural Malawi" *AIDS* 17(17): 2509-2513.

Dupas, P. (2011) "Do teenagers Respond to HIV rRisk Information? Evidecne from a Filed Experiment in Kenya" American Economic Journal: Applied Economics, 3(1): 1-34.

Durevall, D. and A. Lindskog "HIV and Inequality: The Case of Malawi" (2009), Scandinavian Working Papers in Economics (S-WoPEc), No 425.

Durevall, D. and A. Lindskog (2011), "Uncovering the Impact of the HIV Epidemic on Fertility in Sub-Saharan Africa: the Case of Malawi", *Journal of Population Economics*. 24(2): 629-55.

Fink, G., and S. Linnemayr (2008), "HIV, Education, and Fertility: Long-term Evidence from Sub-Saharan Africa", mimeo, Harvard School of Public Health.

Fortson, J. (2009), "HIV/AIDS and Fertility", *American Economic Journal: Applied Economics*, 1(3): 170-194.

Glick, P., and D. Sahn (2008), "Are Africans Practicing Safer Sex? Evidence from Demographic and Health Surveys for Eight Countries", *Economic Development & Cultural Change*, 56(2): 397-439.

Grieser, M., J. Gittelsohn, A.V. Shankar, and T. Koppenhaver (2001), "Reproductive Decision Making and the HIV/AIDS Epidemic in Zimbabwe", *Journal of Southern African Studies* 27(2): 225-243.

Holmqvist, G. (2010) "Fertility impact of social transfers in Sub-Saharan Africa: What about pensions" BWPI working papers 119, Brooks World Poverty Institute, Manchester.

Juhn, C., Kalemli-Ozcan, S., Turan, B. (2009) "HIV and Fertility in Africa: First-Evidence from Population Based Surveys" NBER working paper 12181 (revised).

Kalemli-Ozcan, S. (2003), "A Stochastic Model of Mortality, Fertility and Human Capital investment", *Journal of Development Economics* 70(1): 103-118.

Kalemli-Ozcan, S. (2010), "AIDS, Reversal of the Demographic Transition and Economic Development: Evidence from Africa", NBER Working Paper 12181 (updated).

Kalemli-Ozcan, S. and B. Turan (2010), "HIV and Fertility Revisited", forthcoming, *Journal of Development Economics.*

Lewis, J. C., C. Ronsmans, A. Ezeh, and S. Gregson (2004), "The Population Impact of HIV on Fertility in Sub-Saharan Africa", *AIDS*, 18 (suppl 2): S35-S43.

Lorentzen, P., J. McMillan, and R. Wacziarg (2008), "Death and Development", *Journal of Economic Growth*, 13(2): 81-124.

Magadi M. A., and A. Agwanda (2010). "Investigating the association between HIV/AIDS and recent fertility patterns in Kenya" *Social Science & Medicine*, 71(2), 335-344.

Measure DHS (2010) *Malawi Demographic and Health Survey Dataset 2004,* available at http://www.measuredhs.com/

Moulton, B.R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units", Review of Economics and Statistics, 72, pp. 334-338.

NSO and Macro International (1994) *Malawi Demographic and Health Survey 1992*, Final report, Calverton and Zomba.

NSO and ORC Macro (2001) *Demographic and Health Survey 2000*, Final report, Calverton and Zomba.

NSO and OCR Macro (2005), *Malawi: DHS 2004,* Final report, Calverton and Zomba.

NSO and UNICEF (2008) *Multiple Indicator Survey 2006: Malaw*i, Zomba.

NSO (2010a) *Thematic report on fertility, Population and Housing Census 2008*, Zomba available at http://www.nso.malawi.net/.

NSO (2010b) *Thematic report on spatial distribution and urbanization*, *Population and Housing Census 2008*, Zomba available at http://www.nso.malawi.net/.

Ntozi, J. (2002) "Impact of HIV/AIDS on Fertility in Sub-Saharan Africa" *African Population Studies*, 17(1): 103-124.

Nugent, J.B. (1985) "The Old-Age Security Motive for Fertility" *Population and Development Review,* Vol. 11, No. 1, pp. 75-97

Oster, E. (2005), "Sexually Transmitted Infections, Sexual Behavior and the HIV/AIDS Epidemic", *Quarterly Journal of Economics*, 120 (2): 467-515.

Oster, Emily (2010) "Estimating HIV Prevalence and Incidence in Africa from Mortality Data," *The B.E. Journal of Economic Analysis & Policy* 10(1), Article 80.

Pörtner, C. (2001) "Children as Insurance" *Journal of Population Economics*, 14(1), 119-136

Setel, P. (1995) "The effects of HIV and AIDS on fertility in East and Central Africa" *Health Transition Review*" 5 (Supplement), pp. 179-189.

Schultz, P. (1997), "Demand for Children in Low Income Countries" Chap. 8 in Handbook of Population and Family Economics, Rosenzweig, M. and O. Stark (eds.), Part A, Elsevier, Amsterdam.

Soares R.R. (2005), "Mortality Reductions, Educational Attainment, and Fertility Choice", *American Economic Review* 95(3): 580-601.

Soares R.R. (2006), "The Effect of Longevity on Schooling and Fertility: Evidence from the Brazilian Demographic and Health Survey", *Journal of Population Economics,* 19(1): 71-97.

Ueyama, M., and F. Yamauchi (2009), "Marriage Behavior Response to Prime-age Adult Mortality – Evidence from Malawi", *Demography* 46(1): 43-63.

UNAIDS, (2010) Global Report: UNAIDS Report on the global AIDS epidemic 2010. Geneva. Available at http://www.unaids.org/GlobalReport/Global_report.htm.

UNGASS (2010) "Malawi HIV and AIDS Monitoring and Evaluation Report: 2008-2009"

Country Progress Report, National AIDS Council, Lilongwe.

UNICEF (2010) *Child Mortality Report 2010*, United Nations Inter-agency Group for Child Mortality Estimation, Geneva.

United Nations (2002) HIV/AIDS and Fertility in Sub-Saharan Africa: A Review of the Research Literature" Department of Economic and Social Affairs, UN Secretariat, ESA/P/WP.174.

Young, A. (2005), "The Gift of Dying: The Tragedy of AIDS and the Welfare of Future African Generations", *Quarterly Journal of Economics* 120(2): 423-466.

Young, A. (2007), "In Sorrow to Bring Forth Children: Fertility amidst the Plague of HIV", *Journal of Economic Growth* 12(4): 283-327.

Appendix

Table A1: Individual-level summary statistics

	Women		Men	
	Mean	Std. Err.	Mean	Std. Err.
Births last 5 years	0.974	0.870		
Ideal number of children	4.056	1.345	4.096	1.523
Age 15-19	0.200	0.400	0.200	0.400
Age 20-24	0.236	0.425	0.179	0.384
Age 25-29	0.181	0.385	0.191	0.393
Age 30-34	0.133	0.339	0.144	0.351
Age 35-39	0.099	0.298	0.087	0.283
Age 40-44	0.083	0.276	0.091	0.288
Age 45-49	0.068	0.252	0.052	0.223
Age 50-54			0.054	0.225
Prior births (5 years ago)	2.220	2.611		
Total number of births	3.194	2.740	2.994	3.328
No or incomplete primary education	0.260	0.439	0.117	0.321
Primary education	0.641	0.480	0.678	0.467
Secondary education	0.097	0.296	0.195	0.396
Higher education	0.002	0.043	0.010	0.099
1st wealth quintile	0.203	0.402	0.150	0.357
2nd wealth quintile	0.227	0.419	0.236	0.425
3rd wealth quitile	0.238	0.426	0.256	0.437
4th wealth quintile	0.217	0.412	0.233	0.423
5th wealth quintile	0.115	0.319	0.124	0.330
Catholic	0.779	0.415	0.786	0.410
Central African Presbyterian Church	0.164	0.371	0.175	0.380
Anglican	0.016	0.127	0.017	0.130
Seventh day adventist/baptist	0.061	0.239	0.058	0.234
Other christian	0.365	0.481	0.379	0.485
Muslim	0.161	0.367	0.132	0.339
No religion	0.012	0.107	0.026	0.159
Chewa	0.314	0.464	0.323	0.468
Tumbuka	0.098	0.297	0.098	0.297
Lomwe	0.201	0.400	0.204	0.403
Tonga	0.019	0.135	0.020	0.140
Yao	0.160	0.366	0.134	0.341
Sena	0.031	0.174	0.035	0.185
Nkonde	0.008	0.088	0.010	0.099
Ngoni	0.091	0.287	0.095	0.293
Other ethnicity	0.079	0.269	0.082	0.274

Source: Measure DHS (2010).

Table A2: District-level variables summary statistics (rural areas)

	Obs.	Mean	Std. dev.	Min	Max
Prime-age mortality	27	0.017	0.008	0.007	0.036
Female prime-age mortality	27	0.014	0.007	0.006	0.031
Male prime-age mortality	27	0.019	0.010	0.007	0.041
Pre-HIV fertility	27	0.181	0.021	0.154	0.226

Source: Measure DHS (2010) and data from 1998 Population and Census supplied by National Statistical Office, Zomba.

Table A3: Original and cluster bootstrap-t statistics for selected specifications

Null-hypothesis	Original sample statistic	Distribution of test-statistic from the cluster bootstrap					
		1st percentile	5th percentile	10th percentile	90th percentile	95th percentile	99th percentile
Specification 1. Table 1							
Age 20-24 =0	14.324	-2.477	-1.247	-1.012	1.616	2.255	2.636
Age 25-29 =0	14.349	-2.938	-1.714	-1.197	1.640	1.886	2.795
Age 30-34 =0	11.106	-3.017	-1.841	-1.295	1.699	2.548	2.942
Age 35-39	6.796	-2.297	-1.538	-1.002	1.699	2.389	3.879
Age 40-44	0.738	-2.023	-1.462	-0.970	1.375	1.701	2.814
Age 45-49	-2.026	-2.321	-1.675	-1.414	1.282	1.781	2.616
Prior births	1.915	-2.913	-1.757	-1.381	1.368	1.702	2.436
Primary education	-0.313	-1.930	-1.561	-1.318	0.879	1.032	2.691
Secondary education	-3.611	-2.306	-1.774	-1.602	1.389	2.284	3.758
Higher education	-2.581	-69.488	-56.548	-48.636	4.901	8.222	14.977
2nd wealth quintile	-1.199	-2.418	-1.114	-0.962	1.123	1.589	2.606
3rd wealth quintile	-3.008	-2.559	-1.638	-1.153	1.328	1.732	2.926
4th wealth quintile	-2.227	-2.292	-1.365	-1.079	1.583	2.047	2.998
5th wealth quintile	-6.256	-2.569	-1.464	-1.110	1.535	2.139	5.134
District fertility before HIV	2.460	-21.599	-16.400	-13.800	1.177	2.044	3.325
Adult mortality	1.551	-13.298	-4.459	-3.610	0.949	1.857	3.986
Specification 2. Table 1							
Age 20-24	-3.159	-2.450	-1.476	-1.212	1.133	1.502	1.779
Age 25-29	-1.391	-2.038	-1.402	-0.905	1.003	1.615	2.075
Age 30-34	-2.058	-1.784	-1.293	-0.908	1.196	1.470	2.283
Age 35-39	-1.664	-1.492	-0.939	-0.611	1.340	1.977	2.768
Age 40-44	-1.991	-2.127	-1.178	-0.850	1.245	1.536	2.340
Age 45-49	-2.374	-1.658	-1.128	-0.815	1.311	1.776	2.996
Living children	2.668	-2.196	-1.513	-1.250	0.851	1.188	1.521
Primary education	-0.309	-1.469	-1.240	-1.049	0.738	1.061	1.787
Secondary education	-0.929	-1.741	-1.296	-1.122	0.950	1.253	1.478
Higher education	-1.581	-4.181	-2.438	-1.967	5.975	7.305	9.949
2nd wealth quintile	1.041	-1.935	-1.313	-1.030	1.217	1.351	1.904
3rd wealth quintile	-0.224	-2.163	-1.559	-1.099	0.974	1.410	2.911
4th wealth quintile	1.083	-2.265	-1.701	-1.319	0.964	1.363	2.150
5th wealth quintile	-1.576	-2.167	-1.466	-0.956	1.198	2.421	4.149
District fertility before HIV	-1.262	-2.297	-1.219	-0.732	2.144	2.584	3.346
Adult mortality	-2.398	-3.100	-0.888	-0.592	4.731	5.223	14.110
Specification 3. Table 1							
Age 20-24	-0.349	-1.532	-1.147	-1.038	1.072	1.589	2.123
Age 25-29	0.544	-2.189	-1.653	-1.073	1.363	1.586	2.286
Age 30-34	2.141	-2.177	-1.306	-1.031	0.985	1.298	1.546
Age 35-39	2.779	-2.070	-1.228	-0.870	0.985	1.130	1.606
Age 40-44	1.722	-1.659	-1.224	-0.933	0.988	1.355	2.201

Age 45-49	1.196	-2.454	-1.637	-0.935	1.067	1.168	1.906
Age 50-54	1.289	-2.320	-1.658	-1.178	0.991	1.104	1.581
Living children	7.567	-1.908	-1.370	-1.018	1.130	1.444	1.985
Primary education	0.487	-1.735	-1.361	-1.196	1.068	1.374	1.651
Secondary education	-2.416	-2.342	-1.339	-1.021	0.901	1.140	2.282
Higher education	-1.625	-1.715	-1.070	-0.713	1.335	1.528	3.132
2nd wealth quintile	-2.567	-1.675	-1.326	-0.962	1.111	1.538	2.378
3rd wealth quintile	-2.311	-1.748	-1.365	-0.917	1.148	1.411	1.818
4th wealth quintile	-3.663	-2.344	-1.329	-0.863	1.145	1.442	1.728
5th wealth quintile	-5.861	-1.226	-0.963	-0.783	1.285	1.714	2.427
District fertility before HIV	-1.151	-2.648	-1.821	-1.331	1.297	1.476	2.274
Adult mortality	-1.960	-3.424	-1.896	-1.183	1.352	2.181	3.611
Specification 4, Table 1							
Age 20-24	24.073	-1.515	-1.252	-1.007	1.583	1.880	2.976
Age 25-29	18.840	-1.937	-1.628	-1.270	1.581	1.966	3.043
Age 30-34	14.429	-2.099	-1.308	-1.194	1.488	1.669	2.666
Age 35-39	8.530	-2.083	-1.691	-1.267	1.116	1.533	2.708
Age 40-44	1.915	-2.918	-1.660	-1.298	1.430	2.049	2.258
Age 45-49	-4.311	-1.945	-1.478	-1.214	1.343	1.898	2.545
Prior births	6.427	-3.051	-1.991	-1.590	1.282	2.074	2.498
Primary education	-0.590	-2.099	-1.474	-1.222	1.374	1.603	3.182
Secondary education	-6.278	-2.388	-1.849	-1.579	1.040	1.564	4.019
Higher education	-1.672	-3.278	-2.491	-1.360	1.114	1.769	2.691
2nd wealth quintile	-0.080	-1.706	-1.435	-1.211	1.176	1.755	4.405
3rd wealth quintile	-1.288	-2.531	-1.657	-1.373	1.102	1.431	2.074
4th wealth quintile	-3.159	-2.099	-1.682	-1.227	1.343	1.931	2.252
5th wealth quintile	-5.305	-1.957	-1.374	-1.060	1.697	2.020	3.130
District fertility before HIV	0.651	-2.937	-2.449	-1.855	1.987	3.553	4.458
Adult mortality	-0.563	-4.404	-2.716	-2.168	2.112	2.855	3.902
Specification 5, Table 1							
Age 20-24	14.426	-2.673	-1.340	-1.049	1.584	2.082	2.665
Age 25-29	14.547	-2.827	-1.688	-1.164	1.641	1.957	2.771
Age 30-34	11.353	-2.937	-1.737	-1.197	1.770	2.652	2.925
Age 35-39	6.792	-2.241	-1.550	-0.938	1.673	2.337	3.852
Age 40-44	0.671	-1.981	-1.403	-0.885	1.417	1.690	2.786
Age 45-49	-2.201	-2.251	-1.595	-1.365	1.213	1.805	2.588
Prior births	2.126	-2.812	-1.866	-1.399	1.211	1.708	2.383
Primary education	-0.455	-1.872	-1.454	-1.283	0.834	1.126	2.618
Secondary education	-3.237	-2.233	-1.998	-1.565	1.414	2.147	3.355
Higher education	-2.710	-65.911	-50.962	-47.740	5.415	9.550	15.383
2nd wealth quintile	-1.075	-2.537	-1.120	-0.965	1.135	1.534	2.515
3rd wealth quintile	-2.772	-2.537	-1.755	-1.103	1.346	1.678	2.999
4th wealth quintile	-2.009	-2.661	-1.415	-0.903	1.500	2.085	3.166
5th wealth quintile	-6.479	-2.715	-1.539	-1.103	1.407	2.000	4.908

Adult mortality	-0.969	-2.378	-1.882	-1.525	39.962	43.082	47.031
Specification 1. Table 3							
Female adult mortality	-0.739	-2.605	-1.114	-0.701	3.681	4.297	6.862
Male adult mortality	2.200	-9.164	-4.789	-3.841	0.474	1.126	2.321
Female = male	-1.486	-2.339	-1.124	-0.536	4.282	5.027	8.526
Specification 2. Table 3							
Female adult mortality	-3.500	-2.095	-1.212	-0.854	4.771	7.713	10.074
Male adult mortality	0.124	-6.139	-3.475	-2.547	1.560	2.187	3.266
Female = male	-2.425	-2.213	-1.290	-0.901	5.653	7.736	11.414
Specification 3. Table 3							
Female adult mortality	-0.702	-2.653	-2.006	-1.684	1.717	2.105	3.940
Male adult mortality	-3.210	-3.014	-1.751	-1.402	1.133	1.448	1.740
Female – male adult mortality	0.138	-2.231	-1.770	-1.424	1.498	1.836	3.440
Specification 1. Table 5							
Adult mortality *Age 15-19	3.278	-4.446	-3.357	-2.466	0.622	1.008	1.313
Adult mortality *Age 20-29	1.441	-2.767	-2.428	-1.968	0.773	1.191	2.069
Adult mortality *Age 30-39	0.259	-2.117	-1.421	-0.984	0.950	1.405	2.398
Adult mortality *Age 40-49	0.011	-1.927	-1.342	-0.951	3.005	3.403	4.173
‘Mortality *Age 15-19’ - ‘Mortality*Age 20-29’	2.218	-2.456	-2.001	-1.482	0.524	0.866	1.244
‘Mortality *Age 15-19’ - ‘Mortality *Age 30-39’	3.293	-3.528	-2.312	-1.933	0.516	0.827	1.343
‘Mortality *Age 15-19’ - ‘Mortality *Age 40-49’	2.341	-4.811	-3.710	-2.786	0.779	1.252	1.473
‘Mortality Age 20-29’ - ‘Mortality *Age 30-39’	1.474	-2.421	-1.770	-1.238	0.779	1.076	3.095
‘Mortality *Age 20-29’ - ‘Mortality *Age 40-49’	1.082	-3.034	-2.489	-1.808	0.809	1.046	1.647
‘Mortality *Age 30-39’ - ‘Mortality *Age 40-49’	0.143	-28.284	-1.368	-0.868	1.114	1.423	1.953
Specification 2. Table 5							
Adult mortality *Age 15-19	-2.501	-2.066	-1.426	-1.091	1.634	1.922	2.572
Adult mortality *Age 20-29	-1.471	-2.912	-2.264	-1.736	0.925	1.700	2.118
Adult mortality *Age 30-39	-2.719	-2.612	-1.990	-1.721	1.098	1.836	11.802
Adult mortality *Age 40-49	-2.415	-2.247	-1.460	-1.137	31.420	34.545	39.480
‘Mortality *Age 15-19’ - ‘Mortality*Age 20-29’	-1.649	-2.895	-1.649	-1.206	3.192	3.794	4.756
‘Mortality *Age 15-19’ - ‘Mortality *Age 30-39’	-0.527	-2.676	-1.448	-0.817	2.543	2.975	4.327
‘Mortality *Age 15-19’ -	0.246	-4.783	-3.086	-2.047	1.110	1.558	2.432

'Mortality *Age 40-49'							
'Mortality Age 20-29' - 'Mortality *Age 30-39'	1.347	-2.404	-1.306	-0.724	1.164	1.429	1.807
'Mortality *Age 20-29' - 'Mortality *Age 40-49'	1.630	-5.785	-4.646	-4.443	1.273	1.490	1.958
'Mortality *Age 30-39' - 'Mortality *Age 40-49'	0.680	-66.078	-4.048	-3.446	1.107	1.510	2.377
Specification 3. Table 5							
Adult mortality *Age 15-19	-3.134	-2.706	-1.906	-1.242	1.349	1.771	2.581
Adult mortality *Age 20-29	-1.172	-3.855	-1.840	-1.274	1.401	1.678	2.450
Adult mortality *Age 30-39	-2.189	-3.008	-1.980	-1.346	1.316	1.726	2.052
Adult mortality *Age 40-49	-0.220	-2.902	-1.678	-1.407	0.922	0.978	1.265
Adult mortality *Age 15-19	-2.557	-1.344	-1.082	-0.872	1.155	1.393	2.112
Adult mortality *Age 20-29	-1.099	-1.961	-1.297	-0.994	1.233	1.619	2.267
Adult mortality *Age 30-39	-2.133	-1.922	-0.873	-0.720	1.526	1.633	2.396
Adult mortality *Age 40-49	1.251	-2.685	-1.232	-0.986	1.135	1.363	1.770
'Mortality *Age 15-19' - 'Mortality*Age 20-29'	-0.788	-1.280	-1.027	-0.558	1.414	1.647	2.201
'Mortality *Age 15-19' - 'Mortality *Age 30-39'	-1.884	-1.186	-0.898	-0.779	1.030	1.343	2.411
'Mortality *Age 15-19' - 'Mortality *Age 40-49'	-3.134	-2.706	-1.906	-1.242	1.349	1.771	2.581
'Mortality Age 20-29' - 'Mortality *Age 30-39'	-1.172	-3.855	-1.840	-1.274	1.401	1.678	2.450
'Mortality *Age 20-29' - 'Mortality *Age 40-49'	-2.189	-3.008	-1.980	-1.346	1.316	1.726	2.052
'Mortality *Age 30-39' - 'Mortality *Age 40-49'	-0.220	-2.902	-1.678	-1.407	0.922	0.978	1.265

Paper IV

Does a diversification motive influence children's school entry in the Ethiopian highlands?

Annika Lindskog

Department of Economics
School of Business, Economics and Law
University of Gothenburg

Household-level diversification of human capital investments is investigated. A simple model is developed, followed by an empirical analysis using 2000-2007 data from the rural Amhara region of Ethiopia. Diversification would imply negative siblings' dependency and be more important in more risk-averse households. Hence it is investigated if older siblings' literacy has a more negative (smaller if positive) impact on younger siblings' school entry in more risk-averse households. Results suggest diversification across brothers, but are not statistically strong, and with forces creating positive sibling dependency dominating over diversification.

Keywords: Diversification, Education, Ethiopia, Uncertainty
JEL Codes: I21, D81, D13

1. Introduction

Investment in education is likely to be the most important investment decision made for most people, and also to have uncertain returns, i.e. as opposed to risky investments with a known variance they are truly unknown,. Still, the mplications of uncertainty for such investment have received little attention in economics. Returns to formal education are uncertain, but the alternative, more learning by doing, is not free of uncertainty either. This makes it difficult to come up with hypotheses about how uncertainty and risk aversion affect investments in education.

A common strategy for dealing with risky returns is diversification. In the rural Amhara region of Ethiopia, and in other rural areas of less-developed countries where there is extensive informal insurance, with parents relying on children for old-age support, diversification could mean some investment in formal education – perhaps directed towards employment in the "modern" sector – and some investment in traditional knowledge, acquired through learning by doing in the household and in the field.[1]

Diversification at the household level implies within-household education inequality, and thus, negative sibling-dependency in education (Lilleør, 2008b). If an older sibling has more education, diversification would mean that a younger one should get less, and spend more time acquiring traditional knowledge. Such diversification should relate to risk-aversion, and be stronger in more risk-averse households.

Are investment in education really affected by a household-level diversification motive? Unique data, that in addition to extensive information about children's schooling has information about risk-preferences of the household head, is used to investigate whether sibling-dependency in education was more negative in households with more risk-averse heads in rural Amhara during 2000-2006, as well as whether diversification took place across all siblings or was gender-specific.

[1] In the literature on child-schooling and labour, investment in traditional knowledge is often considered just child work, expanding current consumption possibilities, but without future rewards. There are a few exeptions though, including Bommier and Lambert (2000) and Lilleør (2008a). Rosenzweig and Wolpin (1985) and Grootaert and Kanbur (1995) also demonstrate the potential usefulness of household and farm-specific knowledge in rural areas of less-developed countries.

Annual school entry probabilities of boys and girls age 6 to 16 are estimated. School entry is analyzed since it, more than education decisions at later stages, is likely to be affected by parental preferences rather than child preferences and ability. Annual school entry probabilities are used so that the full sample of children can be used without problems of censoring. To control for time-constant unobserved parental preferences, a linear probability model with household fixed effects is used.

Total siblings' dependency in education turn out to be positive; hence other forces dominate diversification. The results still suggest diversification across brothers; older brothers' education does not have the same positive impact on boys' school entry in households with the most risk-averse heads. However, the diversification results are not statistically strong.

The next section provides a theoretical framework for the study, including the development of the model used. Section 3 then describes and explains the empirical approach, while Section 4 describes the data and variables while giving some background on the study area and on education in Ethiopia. Section 5 presents and discusses the results, and Section 6 summarizes and draws conclusions.

2. Theoretical framework

Economic theory about education is dominated by human capital theory, according to which people invest in education as long as the marginal benefit exceeds marginal costs. Marginal costs include direct costs as well as opportunity costs, while the marginal benefit mainly consists of increased future income (Becker, 1962; Ben-Porath, 1967). Though the literature on determinants of educational investment in less-developed countries is extensive, it has focused on the cost rather than the benefit side. Poverty and credit constraints, high opportunity costs of child time, and supply-side constraints such as lack of nearby schools or sufficient teachers, are generally considered the main factors keeping children out of school (Jacoby, 1994; Edmonds, 2006; Gitter and Barham, 2007; Orazam and King, 2008; Huisman and Smits, 2009).

2.1 Expected returns and uncertainty of returns to educational investment

According to theory, expected returns to education should matter for educational investment decisions, and empirical evidence from economics subfields suggests that they do. In part, returns to education are determined by the child's ability to transform time spent in school into marketable knowledge and skills. As data on test-scores has become more available, studies have been done on the impact of various school inputs on test-scores and on demand for education, and of test scores on school continuation and future labour-market outcomes (Card and Krueger, 1992; Glewwe, 2002; Glewwe and Kremer, 2006; Hanushek, 2008; Hanushek and Woessmann, 2008; Akresh et al., 2010). At least when the family is not too poor, demand for education has also been shown to respond to regional variations in returns to education (Anderson et al., 2003; Kochar, 2004; Kingdon and Leopold, 2008; Chamarbagwala, 2008),[2] as well as to subjective perceptions about returns to education (Attanasio and Kaufmann, 2009; Jensen, 2010).[3]

However, the effect of uncertainty of returns to education has received very little attention. There are many possible sources of uncertainty: about the quality of education, about the child's ability, health and survival throughout adulthood, about future market returns; and about the child's future filial transfers.

Theoretically, risky returns to education can decrease time in school (Lehvari and Weiss, 1974), but the relative riskiness of more versus less education is what matters (Kodde, 1986). Pouliot (2006) introduces uncertainty into Baland and Robinson's (2000) influential model on child labour, and demonstrates that, without perfect insurance markets, uncertainty could lead to inefficiently low levels of education, even with perfect credit markets and parents able to impose filial transfers from children. Estevan and Baland (2007) consider a particular source of uncertainty – young adult mortality – and again that, without perfect insurance markets, uncertainty could lead to inefficiently low levels of education when parents want filial transfers rather than planning on parental transfers (bequests) to their children.

[2] But Nerman and Owens (2010) find that returns do not determine demand in Tanzania.
[3] That schooling of children in poor households responds less to differences in expected returns than does that of children in richer households is suggestive of the importance of credit constraints and poverty.

There is little empirical evidence on the importance of uncertainty for educational investments, perhaps because uncertainty is impossible to observe and measure. In an Italian sample, Belzil and Leonardi (2007) find that educational investment is negatively related to risk-aversion, with a small but statistically significant effect. Attanasio and Kaufmann (2009) find that perceived subjective employment and wage risks affect the decision to continue into senior high school in urban Mexico, though again the effect is small.

As mentioned earlier, it is not obvious that investments in formal education are (perceived to be) more uncertain than the relevant alternative. In rural Ethiopia the alternative to formal education is learning by doing in the household and on the farm, which is investment in traditional knowledge. But rain-feed agriculture in the Ethiopian highlands is definitely not free of uncertainty. Furthermore, it is possible to view education as insurance, making the individual better able to manage in an uncertain future.[4]

Still, education is more likely to lead to migration and urban employment, with higher probability of unemployment, and less parental control of children. Empirical evidence on returns to education and on unemployment probabilities in Ethiopia is scarce and not always consistent, but to have more than a couple of years of education appears to yield high returns in cities, though not in rural areas – while unemployment rates are higher for the better educated (World Bank, 2005).

Lack of experience with education and with non-agricultural employment in rural Ethiopia may also create subjective uncertainty about the returns expected from education, which ought to be what matters for actual decisions. In the Dominican Republic teenagers living in neighborhoods with few well-educated were found to underestimate returns to education, while provision of correct information increased their schooling (Jensen, 2010). Since returns to both formal education and traditional knowledge are uncertain it is hard to hypothesise about the effect of uncertainty or of risk aversion on the level of education.

[4] The idea that education has a return primarily during times of change has been around since Schultz (1975) and Foster and Rosenzweig (1996).

2.2 Diversification of human capital investment

Independent of which investment is viewed as most uncertain – formal education or traditional knowledge – diversification is a possible strategy. To some extent, it is possible to diversify at the individual level by providing a child with both formal education and traditional knowledge. A reason for delayed school entry in rural areas of less-developed countries can be a desire that children should first gain some basic traditional knowledge (Bommier and Lambert, 2000).

However, scope for diversification is probably larger at the household level, where, for example, one sibling can get more formal education while the others spend more time acquiring traditional knowledge. Household-level diversification has been proposed to matter for rural-urban migration, an issue naturally connected investment in education (Levhari and Stark, 1982).

Lilleør (2008a; 2008b, 2008c) argues that household-level diversification of educational investment should be especially important where people rely on mutual support within the extended family for insurance and old-age support. Such diversification can result in negative sibling-dependency in education, i.e. more education of older siblings being negatively related to younger sibling's education. Using Tanzanian data, Lilleør's (2008b) finds negative sibling-dependency among sons – who are perhaps more likely to support parents when old - when a large share of older brothers is well-educated (the total effect is non-linear; when older brothers have less education the effect is positive).

A problem when attempting to investigate diversification by analysing sibling-dependency in education is that many other factors – e.g. credit constraints, unobserved parental preferences, or positive within-household education spillovers – could also affect siblings' dependency[5]. However, there should be more diversification in more risk-averse households, while there is no reason to expect risk aversion to matter for the other factors creating sibling-dependency. Thus, we can test for diversification as a motive influencing children's school entry by analysing how sibling-dependency differs in differently risk-averse households.

[5] In the fifth thesis chapter total sibling-dependency in education and what explains it is analyzed, using the same data from rural Amhara region of Ethiopia. Thus, a more detailed description of other mechanisms that can create sibling-dependency is found there.

2.3 A simple model of diversification

The purpose of this model is to illustrate the motivation to diversify, and how this differs with risk aversion, not to offer a complete and realistic model of what determines children's school entry or total schooling. For now, let's therefore completely abstract from inter-temporal aspects of the human-capital investment decision, making essentially a one-period model, even though we will consider the impact of current human-capital investment on expected future income and consumption. Since savings/debts, as well as current-period costs and gains from formal education or from traditional knowledge, i.e. the fruits of labour, are left out of the model, credit constraints are not an issue.

The household consists of one parent, who is the decision-maker, and two children, one older and one younger. Parents get utility from consumption, and have a concave utility function, meaning that they are at least to some degree risk averse.

Children's time can be allocated to formal education (ed) or learning of traditional knowledge (tk), so that $tk^{old} = 1 - ed^{old}$ and $tk^{young} = 1 - ed^{young}$. Parent's future consumption depends on the random return to children's time invested in formal education (R^{ed}) and traditional knowledge (R^{tk}) where expectations on older and younger siblings' contributions are the same, conditional on their human capital; $E\left[R^{tk\_young}\right] = E\left[R^{tk\_old}\right] = \mu^{tk}$ and $E\left[R^{ed\_young}\right] = E\left[R^{ed\_old}\right] = \mu^{ed}$. We assume that human capital investment of the older child's time has already been made, leaving only the decision about the younger child's human capital investment.

The amount of formal education is chosen to maximize the expected utility of future consumption, $E[u(c)] = E\left[u\left(R^{tk}\left(2 - ed^{old} - ed^{young}\right) + R^{ed}\left(ed^{old} + ed^{young}\right)\right)\right]$, resulting in the following first order conditions:

$$\begin{array}{lll} E\left[u'(c)R^{ed}\right] = E\left[u'(c)R^{tk}\right] & & ed^{young} \in (0,1) \\ E\left[u'(c)R^{ed}\right] > E\left[u'(c)R^{tk}\right] & & ed^{young} = 1 \\ E\left[u'(c)R^{ed}\right] < E\left[u'(c)R^{tk}\right] & & ed^{young} = 0 \end{array} \qquad (1)$$

The expected marginal utilities should thus be the same from investment in formal education and in traditional knowledge. At the extremes, when they differ

"sufficiently", the younger child will specialize completely in either formal education or traditional knowledge.

Focusing on the interior solution, the first-order condition can be rewritten

$$E[u'(c)]\mu^{ed} + Cov(u'(c), R^{ed}) = E[u'(c)]\mu^{tk} + Cov(u'(c), R^{tk}) \quad (2)$$

where the covariance terms are negative and dependent on the total amount invested in formal education, $ed^{old} + ed^{young}$. The diversification motive stems from this dependence of the covariance terms on the total level of investment in formal education. With more formal education, consumption depends more on the return to formal education and less on the return to traditional knowledge, making the right-hand side larger and the left-hand side smaller. Thus, if the expected return to formal education is not too different from that to traditional knowledge, parents will want to diversify and make consumption dependent on both types of human capital, rather than just one. When the older sibling has more formal education, the optimal level of formal education for the younger child consequently becomes smaller.

Now let's assume that $v(c)$ is the utility function of a more risk averse person than the person with utility function $u(c)$. This means that $v(c)$ is more concave than $u(c)$; formally, there is an increasing concave function $\psi(\cdot)$ such that $v(c) = \psi(u(c))$. This makes $v(c)$ more sensitive to dispersions of consumption around the expected value.[6] Thus, everything else equal, the absolute values of the covariance terms $Cov(v'(c), R^{ed})$ and $Cov(v'(c), R^{tk})$ are larger than the absolute values of $Cov(u'(c), R^{ed})$ and $Cov(u'(c), R^{tk})$. Moreover, if $u(c)$ is replaced by $v(c)$, in absolute terms, the larger of the two covariance terms, $Cov(u'(c), R^{ed})$ and $Cov(u'(c), R^{tk})$, will increase more than the smaller one. Everything else equal, the larger covariance term will, by equation (2), be the covariance term pertaining to returns to investments in the type of human capital with the highest expected return, and therefore where most investment should be made. Thus, an increase in risk-aversion will imply a shift from investments

[6] How sensitive $v(c)$ and $u(c)$ are to differences in expected returns depends on the specific functional forms. Sometimes $v(c)$ will be more sensitive at 'sufficiently' low levels of consumption. At 'sufficiently' high levels of consumption (sometimes always), $v(c)$ will be less sensitive.

in the type of human capital with the highest expected return, towards a more diversified combination.

3. Empirical approach

Sibling-dependency in education is investigated by estimating the effect of older siblings' literacy on the annual conditional school entry probability. The first sub-section describes and motivates the empirical model. The second sub-section brings up the possibility of sample-selection bias.

3.1 The annual school entry probability

While parents are almost invariably assumed in economics to be the sole household decision-makers, it is quite likely that children themselves influence education decisions more and more as they grow older and/or gain more school experience.[7] But, as noted earlier, school entry should be more influenced by parents' preferences, on risk for example, than by the child's preferences. School entry should also be less influenced by the child's relieved ability in school and 'taste' for school.

There are practical advantages with focusing on school entry, which ensures a fairly large sample with a fair degree of variation in the dependent variable. All children over a certain age have been eligible for school entry at some point (though not necessarily between 2000 and 2006), and information about ever having attended school, and school-entry age, is available for most children in the data.

The official school-entry age in Ethiopia is 7, but some enter at 6, and many enter later, particularly in rural areas. In the empirical analysis, a child is classified as eligible to enter school if between 6 and 16 years old and never attended before. If a children eligible to enter school does not, this means either that they will never enter or that they will enter later. Never attending school would mean complete specialization in traditional knowledge, while, as noted earlier, delayed school entry

[7] In urban Mexico, preferences of both the child and the mother have been found to matter for the decision to continue into senior secondary school, while only preferences of the child matter for the decision to start college (Attanasio and Kaufmann, 2009). It thus seems reasonable that preferences of the parents are most important for initial school entry.

may be the result of a desire that the child first acquire some basic traditional knowledge (Bommier and Lambert, 2000).

The annual conditional school entry probability is essentially a discrete-time "hazard" model. The advantage of modelling it instead of duration until school entry (or age at school entry) is that the data can be fully used while avoiding censoring problems. There is no need to restrict the sample to children old enough to know that most of those who haven't entered will not do so later. This is especially useful for the results to remain relevant in a situation, such as Ethiopia, where schooling has increased massively in recent years.

The main explanatory variable used here is literacy of older siblings, which is clearly endogenous to parental characteristics affecting education decisions regarding all children in the family. Some of these characteristics, such as parental attitudes towards formal education versus investment in traditional knowledge, and towards child human capital (education and health) investment in general, are unobservable. Unobserved household effects must therefore be controlled for. Doing so with random effects would be problematic, since unobserved parental characteristics are likely to be correlated with older siblings' literacy. For this reason, household fixed effects are used.

There could still be problems of time-varying shocks to the household, affecting education both of the child and of older siblings. Older siblings' education is lagged, which would deal with time-varying household effects in a model without fixed effects. But with fixed effects, strict exogeneity is a must; that is, the explanatory variables must be uncorrelated with lags and leads of the error term (Arellano, 2003). One way to deal with time-varying shocks is to include measures of self-reported health and environmental shocks as done here. As opposed to variables measuring income or wealth over time, these should be exogenous to older siblings' education.

To be able to include household fixed effects, a linear probability model is used, though the common procedure would be to estimate a conditional probability with a logit model. But using household dummies in a non-linear model, such as the logit, leads to biased estimates when not many observations per household are available (Lancester, 2000). A conditional fixed-effects logit model could be an alternative, but

then observations' from households without variation in the dependent variable would not be used (Hsiao, 2003). The disregarded observations would be from households where no child enters school or where all children enter as soon as they become eligible, and thus disregarded households would be special with regards to preference for diversification. A disadvantage with the linear probability model is that we may end up with predicted probabilities below 0 or above 1.

The conditional school-entry probability of child *i* in household *h* and year *t* is

$$\Pr(y_{it}=1)=\beta_0+\sum_{k=1}^{3}\beta_k lit_{it}^{old} risk_h^k+\beta_4 sib_{it}+\beta_5 age_{it}+\beta_6 z_{ht}+\alpha_h+\tau_t+\varepsilon_{it} \quad (3)$$

which depends on the education of older siblings measured by their literacy (lit_{it}^{old}), with the effect allowed to differ depending on three categories of risk preference of the household head $(risk_h^k)$, with $k=1$ being the least risk averse and $k=3$ the most. The school entry probability also depends on: the number of older siblings, (sib_{it}) capturing effects of birth-order rather than of the total number of siblings in a within-household model; child age (age_{it}); self-reported shocks to the household (z_{ht}); time-constant household characteristics (α_h) such as parental education, permanent income and unobserved parental attitudes towards education; year effects, (τ_t) capturing both the massive expansion in primary education in Ethiopia over the study period and possible effects of aging of the households in the panel; and an error term (ε_{it}).

The main hypothesis to be tested is $\beta_1 > \beta_2 > \beta_3$, i.e. that the effect of older siblings' literacy is more negative (smaller if positive, larger if negative) in more risk-averse households. If the diversification motive is strong enough to dominate factors suggesting positive sibling-dependency in education, we should also find $\beta_1, \beta_2, \beta_3 < 0$.

Estimations are done separately for boys and girls since it is quite possible that diversification is gender-specific, that it takes place across brothers or across sisters, rather than across siblings in general. For example, if older brothers have little formal education (and much traditional knowledge), parents might want the younger brother

to have more formal education, and hence that he enters school early, but their preferences on the younger sister's education might not be affected.

3.2 Is there sample selection bias?

Since we are interested in the effects of older siblings' literacy, obviously only children with older siblings can be included in the estimations. So first-born children were excluded, when estimating gender specific effects both first-born sisters and first-born brothers. Smaller families can then be disproportionally excluded, which might be problematic as they might differ from larger families in important ways.

According to the child quantity-quality trade-off hypothesis, parents will beforehand make decisions about the number of children and how much to invest in their education and health (Becker and Lewis, 1973). More education-friendly parents may choose to have fewer children in order to invest more in each. Moreover, more risk-averse parents may choose to have more children (Cain, 1983).

To investigate possible sample-selection bias, excluded households are compared with all others in Table 1, with regard to educational indicators, risk aversion of the household head, and some other household characteristics.

Table 1: Comparison of included households and those excluded since they only had first-born children in the sample

	Excluded households		Included households		Equal means test	
	Mean	St. dev.	Mean	St. dev.	t-stats	p-value
Entry rate	0.31	0.26	0.28	0.25	0.94	0.335
Grade-progress rate	0.95	0.10	0.94	0.14	1.03	0.302
Literacy of household head	0.37	0.48	0.46	0.50	-1.69	0.091
Literacy of spouse	0.22	0.41	0.21	0.41	0.15	0.882
Household size	3.49	2.45	5.57	1.95	-7.78	0.00
Age of household head	48.03	20.76	46.50	14.08	0.59	0.552
Most riskaverse head	0.18	0.38	0.17	0.37	0.33	0.742
Middle riskaverse head	0.47	0.50	0.50	0.50	-0.65	0.515
Least riskaverse head	0.36	0.48	0.34	0.47	0.42	0.671

Excluded households tend to be much smaller, and are less likely to have literate heads (if the smaller size had anything to do with a quantity-quality trade-off, we would have expected smaller households to have more literate heads). Excluded

households do not differ much from others with regard to children's school-entry and -progress rates,[8] or with regard to risk aversion.

4. The data and variables

The data used here comes from the Ethiopian Environmental Household Survey (EEHS), collected by the Ethiopian Development Research Institute (EDRI) in cooperation with the University of Gothenburg and, during the last round, with the World Bank. Four rounds of data have been collected, in 2000, 2002, 2005, and 2007. Interviews were conducted in April/May, towards the end of the Ethiopian school year, which starts in September and ends in June.

The sampled households were from 13 Kebeles in the South Wollo and East Gojjam zones of the Amhara region. The two zones were chosen to represent different agro-climatic zones in the Ethiopian highlands: There is less rainfall in South Wollo than in East Gojjam. Most households in the study areas make their living from rain-fed subsistence agriculture. Access to roads and to capital markets is quite limited.

Two of the Kebeles were added in the third round in order to evaluate a land certification program. The other eleven Kebeles were chosen randomly within the two zones. Within each 120 households were randomly selected. On average an interview took 1.6 days to complete. When a household was not located in a follow-up survey, it was replaced with another, randomly selected, household.

Most of the information on children's education was collected in the fourth round, when respondents were asked about the schooling history of all household members aged 6 to 24. It was attempted to collect data for household members no longer residing in the household, but less successfully, resulting in more missing and incomplete data for non-resident household members. Data from the fourth round, is her to create an annual panel on entry into first grade.

[8] But consistent with the empirical evidence from many less-developed countries, oldest siblings (excluded from the estimations) generally have rates than did their younger siblings (results not reported).

In the fourth-round sample there are 5,160 children aged 6 to 16 from 1,652 households. A requirement that the household be present in at least one previous round, and that there be relatively stable risk preferences of the household head over the third and fourth rounds, reduces the sample to 3,694 children from 1,171 households.[9] Excluding children with no older sibling further reduces the sample to 2,402 children from 875 households (using only children with both an older brother and an older sister, as is done in some estimations, reduces it to 1,766 children from 638 households). Of the remaining children, school entry data is available for 94.6%; 21.1% have never attended school and 73.5% have information about age at school entry. To be included in the estimation children have to be eligible for school entry at some point during 2000-2006 and information on explanatory variables has to be available. This leaves 1,094 children from 527 households in the final main sample.

A central explanatory variable is the household head's risk preference. In the third and fourth rounds the household head did risk aversion experiments, being asked to make pair-wise choices between plots that differed with respect to their yields in good times and bad, with a 50/50 probability of each. One plot had a higher expected yield, but a lower certain (bad times) yield. Based on a sequence of choices, each household was given a risk-preference rank of 1 to 5.

Risk preferences expressed at a specific time are likely to have both a time-constant part, i.e., underlying exogenous preference, and a context-dependent part, which might vary with income and wealth for example. Here focus is on the time-constant part. The mean from the two rounds was calculated and three dummies created for differently risk-averse heads.[10] To increase the reliability of the data households where risk preferences changed too much between the two rounds (about one third of the sample) were also excluded. This probably also eliminate households that experienced large income shocks between the rounds.[11]

[9] Including also households with less stable risk preferences don't qualitatively affect the results.

[10] An alternative is to compute a risk-aversion parameter and interact this with the literacy of older siblings. This doesn't qualitative change the results (results not reported). When computing the risk-aversion parameter, constant relative risk aversion was assumed. The mid-points of ranges were used for the three middle ranks and the least extreme end-points for the most and least risk-averse ranks. Means over the two rounds were then calculated.

[11] Using the same data, Damon et al., (2011) considered determinants of changed time preferences between rounds, and find environmental shocks to be the major determinant.

The data on the number of older siblings is from the last preceding round and includes both those living in the household and any who might have left. As noted, siblings' education is measured by their literacy rate. Though a rough indicator of investment in education, literacy has the advantage of few missing values.

To control for time-varying shocks to household income, dummies indicating the self-reported occurrence of health and environmental shocks are used. Health shocks are either the death or serious illness of a household member. Environmental shocks are mainly draughts and floods, but also other weather-related shocks as well as pests affecting plants or animals. Each dummy was set to one if the shock had occurred at least once between rounds.[12]

An age control is also included and possible gender effects are distinguished by running separate estimations for boys and girls. Summary statistics for all variables included in the model are reported in table A1 in the appendix.

5. Education in Ethiopia and among children in the data

There have been dramatic changes in primary education in Ethiopia recently, with massive increases in enrolment, albeit from a very low starting point. The changes started with the 1994 Education Reform, followed, so far, by three Education Sector Development Programs. The reform in 1994 abolished school fees, and since then decision-making has been decentralized and community involvement in schools has been encouraged. Moreover, many new schools have been built: The number of primary schools increased about 50% during 2000-2004, with the largest increase in rural areas. The budget share for education has also increased, from 13.8% in fiscal year 2000/01 to 19% in 2004/05. As a result, enrolment rates have steadily increased at all stages of education: The gross primary school enrolment rate rose from 34.0% in 1994/95 to 91.3% in 2005/2006, and net enrolment from 36.0% in 1999/2000 to

[12] If there was a shock during 2005-2006, the dummy was set to equal one for both these years. While there is information about the timing of the last shock in the data, there is no information about the timing of earlier shocks, so an annual shocks series could not be created. Estimations using an index of wealth in the preceding round instead of shocks where also run, which did not qualitative affect results. Since income and wealth could be endogenous to older siblings' education, the shocks variables were preferred despite their limitations.

77.5% in 2006/07.[13] Furthermore, the gender gap has been narrowed; the gender parity index increased from 0.6 in 1997/98 to 0.84 in 2005/2006. As is common with such large expansions in enrolment, the numbers of teachers and classrooms have not increased at pace with the number of pupils, raising concerns about reduced quality (Oumer, 2009; Ministry of Education, 2005; World Bank, 2005).

Ethiopia is a large and diverse country, and there are large regional variations in gross and net enrolment rates, as well as in gender disparities. In Amhara net enrolment in years 2004/2005 was 54.6% for boys and 53.1% for girls, both lower than the country averages, but Amhara is one of the few regions where net enrolment appears to be nearly as high for girls as for boys (Ministry of Education, 2005).

Using the school-entry data collected in the fourth round, Table 2 reports the shares of 8 and 11 year old children who had started school over time. It was common to start late, the share who has started by age 8 is around 20% in the mid-1990s and approach 60% after the mid-2000s, while a larger share of 11-year olds have started (around 30% in the mid-1990s and approaching 85% after the mid-2000s). Still in 2006, many children appear to never start school at all, or at least they had not yet done so by age 11.

Table 2: Share of 8 and 11 years old boys and girls that had started school over time

Year	Girls age 8	Boys age 8	Girls age 11	Boys age 11
1996	0.20	0.23	0.29	0.30
1997	0.22	0.13	0.40	0.35
1998	0.30	0.23	0.41	0.47
1999	0.32	0.28	0.54	0.51
2000	0.38	0.31	0.60	0.54
2001	0.45	0.41	0.65	0.51
2002	0.46	0.44	0.74	0.62
2003	0.43	0.45	0.69	0.63
2004	0.53	0.53	0.84	0.71
2005	0.56	0.58	0.80	0.75
2006	0.61	0.59	0.82	0.80
2007	0.56	0.35	0.84	0.86

Information on child age and if the child has yet started school is from the spring in the relevant year. Thus, a child that have not started could start in the autumn that year.

[13] The gross primary school enrolment rate is the ratio of number of pupils enrolled in primary school to the number of children in primary-school age. The net primary school enrolment rate is the ratio of the number of pupils in primary-school age enrolled in primary school to the total number of children in primary-school age.

6. Regression analysis

Estimations of annual school entry probabilities were run for girls and boys separately, treating all older siblings the same (columns 1 and 3, Table 3) and distinguishing sisters from brothers (columns 2 and 4, Table 3).

Table 3: Coefficents from linear estimations with household fixed effects of the effect of diversification on the annual school entry probability.

	Girls		Boys	
	(1)	(2)	(3)	(4)
Number of older siblings	-0.002		0.014	
	(0.012)		(0.011)	
Older siblings' literacy rate *low risk-aversion	0.123*		0.122	
	(0.069)		(0.079)	
Older siblings' literacy rate *middle risk-aversion	0.040		0.231***	
	(0.082)		(0.069)	
Older siblings' literacy rate *high risk-aversion	0.098		0.072	
	(0.147)		(0.096)	
Number of older sisters		-0.005		0.001
		(0.022)		(0.024)
Older sisters literacy rate *low risk-aversion		0.046		0.038
		(0.098)		(0.080)
Older sisters literacy rate *middle risk-aversion		0.156*		0.149*
		(0.090)		(0.085)
Older sisters literacy rate *high risk-aversion		0.042		0.140
		(0.063)		(0.134)
Number of older brothers		-0.001		0.073***
		(0.027)		(0.027)
Older brothers literacy rate*racy rate low risk-aversion		0.089		0.221**
		(0.063)		(0.086)
Older brothers literacy rate*middle risk-aversion		0.023		0.223**
		(0.080)		(0.098)
Older brothers literacy rate*high risk-aversion		0.004		-0.007
		(0.149)		(0.172)
Age	0.011**	0.009	0.026***	0.035***
	(0.005)	(0.007)	(0.005)	(0.008)
Year 2001	0.069**	0.077**	0.055**	0.083**
	(0.027)	(0.033)	(0.024)	(0.033)
Year 2002	0.137***	0.156***	0.095***	0.091**
	(0.033)	(0.039)	(0.025)	(0.036)
Year 2003	0.194***	0.237***	0.144***	0.143***
	(0.035)	(0.045)	(0.030)	(0.041)
Year 2004	0.220***	0.252***	0.145***	0.103*
	(0.044)	(0.057)	(0.038)	(0.056)
Year 2005	0.250***	0.275***	0.219***	0.201***
	(0.047)	(0.059)	(0.041)	(0.060)
Year 2006	0.276***	0.294***	0.155***	0.089
	(0.054)	(0.067)	(0.045)	(0.069)

Table 3 cont.

Health shock	0.027	0.027	-0.031	-0.049
	(0.041)	(0.052)	(0.039)	(0.050)
Environmental shock	-0.017	0.022	0.007	0.043
	(0.031)	(0.039)	(0.035)	(0.047)
Constant	-0.064	-0.105	-0.275***	-0.554***
	(0.073)	(0.105)	(0.071)	(0.139)
Observations	2194	1337	2553	1361
Children	693	413	794	420
Households	465	274	517	288

Standard errors, clustered at the household, in parenthesis.
*= p<0.10, **=p<0.05, ***=p<0.01

Older siblings' literacy seems to matter for school-entry probabilities, with a positive effect. In all four estimations there are statistically significant large positive effects for at least one risk-aversion category. Thus, diversification motive is not strong enough to dominate forces creating positive sibling-dependency. But in line with a diversification motive, the positive effects are found in low and middle, not high, risk-aversion households. For boys there is a statistically significant 22 percentage point increase in the school entry probability if older siblings are literate rather than illiterate, when the head has middle risk aversion. When the household head has low risk aversion the probability of school-entry increases with about 12 percentage points for girls and boys alike. For girls the effect is just significant at the ten percent level and for boys it is just not. Effects in high risk-aversion households, and effects on girls' school entry in middle risk-aversion households, are far from statistical significance. However, coefficients do not follow the hypothesized pattern $\beta_1 > \beta_2 > \beta_3$. For boys the effect is larger in mid- than low-risk aversion households. The null hypothesis of equal effects of older siblings' literacy in differently risk averse households cannot be rejected (Table 4).

Table 4: Is the effect of older siblings' literacy equal across levels of household risk-averseion (F-tests on equality of coefficients)?

	Boys		Girls	
Null hypothesis	F-statistic	p-value	F-statistic	p-value
Literacy of older siblings has an equal coefficient across levels of household risk aversion	1.05	0.351	0.31	0.731
Literacy of older sisters has an equal coefficient across levels of household risk aversion	0.53	0.588	0.64	0.530
Literacy of older brothers has an equal coefficient across levels of household risk aversion	0.73	0.481	0.28	0.754

Distinguishing older sisters from older brothers (columns 2 and 4, Table 3) gives results suggestive of diversification across brothers. In households with not too risk-averse heads, the school entry probability for boys is about 22 percentage point higher when older brothers are literate, while the effect is small and not statistically significant in households with the most risk-averse heads. Diversification across brothers, but not across sisters, is consistent with the findings of Lilleør (2008b). Again, the null hypotheses of equal effect of older sisters' and brothers' literacy in differently risk-averse households cannot be rejected (Table 4).

The number of older siblings had little effect on the annual school-entry probability (Table3). The exception is a positive effect for boys of having more older brothers; it might thus be beneficial to be later born among the boys. This says nothing about any adverse effect of being first-born, or being the first-born sister, often discussed in the literature, since first-born are not in the sample.[14]

Age increases school entry probability more for boys than for girls. This is consistent with boys being more likely to enter late. The omitted reference year is 2000; the coefficients of the year dummies reveal that, over time, girls' school-entry probability increased more than boys'. There are no statistically significant effects of health shocks or environmental shocks.

7. Summary and conclusions

Household-level diversification in human capital investment in a rural area in a less-developed country was investigated. Returns both to formal education and to investment in traditional knowledge are uncertain. A simple model shows that unless expected returns differ too much, parents therefore want to diversify, resulting in negative siblings' dependency in education. If older siblings have less formal education and more traditional knowledge, the young sibling will have more formal education. The desire for diversification should be stronger when the household head is more risk-averse, implying that sibling-dependency in education should be more strongly negative as well.

[14] Just comparing means, first-born children indeed appear to get less education than others.

The empirical analysis estimated the effect of older siblings' literacy on annual school entry probability, using data from rural Amhara region in Ethiopia. Rural Amhara is a place with extensive informal insurance and where parents are likely to depend on children in old age, hence a place where household-level diversification could be of importance. The effect of older siblings' literacy was allowed to differ across households with differently risk-averse heads. Estimations were done separately for girls and boys, and distinguishing effects of older sisters from effects of older brothers.

Total sibling-dependency in education was found to be positive, so any diversification was dominated by other forces. But in line with diversification across brothers, the effect of older brothers' literacy was smaller (more negative) in households with the most risk-averse heads. Having literate rather than illiterate older brothers was found to increase the annual school-entry probability with 22 percentage points in households where the head was not too risk-averse. In households with the most risk-averse heads there is no effect of older brothers' literacy, i.e. the effect of having literate older brothers was estimated to be around 22 percentage points lower. However, the coefficient of older brothers' literacy in the households with the most risk-averse heads, in addition to being close to zero, had a large standard error, so that the null hypothesis of an equal effect in these households could not be rejected. For girls no influence of diversification could be detected.

Summing up, diversification across brothers possibly influenced boys' school entry in rural Amhara in 2000-2006. Possible diversification across brothers, but not across sisters, has been found also in rural Tanzania (Lilleør, 2008b). However, the 'diversification across brothers' result in this paper was statistically weak, and any diversification was dominated by factors creating positive sibling-dependency in education.

Considering the very limited research on diversification of human capital investment, more research is certainly needed. The issue has important policy implications. If a desire for diversification is important to rural households, policies that makes it easier to combine school with farm work and household work – or that makes the curriculum more relevant for such traditional activities in rural areas – could result in more schooling for children from rural households with risk-averse decision-makers.

References

Akresh, R., E. Bagby, D. de Walque, H. Kazianga (2010), "Child Ability and Household Human Capital Investment Decisions in Burkina Faso", IZA Discussion Paper 5326.

Anderson, K.H., E.M. King, Y. Wang (2003), "Market Returns, Transfers and Demand for Schooling in Malaysia, 1976-89", *Journal of Development Studies* 39(3): 1-28.

Arellano, M. (2003), "Panel Data Econometrics", Oxford University Press, New York.

Attanasio, O., K. Kaufmann (2009), "Educational Choices, Subjective Expectations, and Credit Constraints", NBER Working Paper 15087.

Baland, J.M., J.A. Robinson (2000), "Is Child Labor Inefficient", *Journal of Political Economy* 108(4): 663-679.

Becker, G. (1962), "Investment in Human Capital: A Theoretical Analysis", *Journal of Political Economy* 70(5): 9-49.

Becker, G., H. Lewis (1973), "On the Interaction between the Quantity and Quality of Children", *Journal of Political Economy* 81(2): S279-S288.

Belzil, C., M. Leonardi (2007), "Can Risk Aversion Explain Schooling Attainments? Evidence from Italy", *Labor Economics* 14(6): 957-970.

Ben-Porath, Y. (1967), "The Production of Human Capital and the Life Cycle of Earnings", The Journal of Political Economy 75(4): 352-365.

Bommier, A., S. Lambert (2000), "Education Demand and Age at School Enrolment in Tanzania", *Journal of Human Resources* 35(1): 177-203.

Cain, M. (1983), "Fertility as an Adjustment to Risk", *Population and Development Review* 9(4): 688-302.

Card, D., A. Krueger (1992), "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools", *Journal of Political Economy* 100(1): 1-40.

Chamarbagwala, R. (2008), "Regional Returns to Education, Child Labour and Schooling in India", *The Journal of Development Studies* 44(2): 233-257.

Damon, M., Difalco, S., Köhlin, G. (2011), "Environmental Shocks and Rates of Time Preference: A Behavioral Dimension of Poverty Traps", mimeo.

Edmonds, E. (2006), "Child Labor and Schooling Responses to Anticipated Income in South Africa", *Journal of Development Economics* 81(2): 386-414.

Estevan, F., J.M Baland (2007), "Mortality risks, education and child labor", *Journal of Development Economics* 84(1): 118-137.

Foster, A. , M. Rosenzweig (1996), "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution", *American Economic Review* 86(4): 931-953.

Gitter, S.R., B.L. Barham (2007), "Credit, Natural Disasters, Coffee, and Educational Attainment in Rural Honduras", *World Development* 35(3): 498-511.

Glewwe, P. (2002), "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes", *Journal of Economic Literature* 40(2): 436-482.

Glewwe, P., M. Kremer (2006), "Schools, Teachers, and Education Outcomes in Developing Countries", Chapter 16 in Hanuschek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, North-Holland, Amsterdam.

Grootaert, C., R. Kanbur (1995), "Child Labour: an Economic Perspective", *International Labour Review* 134(2): 187-203.

Hanushek, E. (2008), "Do Students Care about School Quality? Determinants of Dropout Behavior in Developing Countries", *Journal of Human Capital* 2(1): 69–105.

Hanushek, E., L. Woessmann (2008), "The Role of Cognitive Skills in Economic Development", Journal of Economic Literature 46(3): 607–668.

Huisman, J., J. Smits (2009), "Effects of Household- and District-Level Factors on Primary School Enrollment in 30 Developing Countries", *World Development* 37(1): 179-193.

Hsiao, C. (2003), Analysis of Panel Data, Second Edition (Econometric Society Monographs No. 34), Cambridge University Press, Cambridge.

Jacoby, H. (1994), "Borrowing Constraints and Progress Through School: Evidence from Peru", *Review of Economics and Statistics* 176(1): 151-160.

Jensen, R. (2010), "The (Perceived) Returns to Education and the Demand for Schooling", *Quarterly Journal of Economics* 125(2): 515-548.

Kingdon, G., N. Leopold (2008) "Do Returns to Education Matter to Schooling Participation? Evidence from India", *Education Economics*, 16(4), 329-350.

Kochar, A. (2004) "Urban Influences on Rural Schooling in India", *Journal of Development Economics* 74(1): 113-136.

Kodde, D. (1986), "Uncertainty and the Demand for Education", *The Review of Economics and Statistics* 68(3): 460-467.

Lancester, T. (2000), "The Incidental Parameters Problem Since 1948", *Journal of Econometric Literature* 95: 391-413.

Levhari, D., O. Stark (1982), "On Migration and Risk in LDCs", *Economic Development and Cultural Change* 31(1): 191-196.

Levhari, D., Y. Weiss (1974), "The Effect of Risk on the Investment in Human Capital", *The American Economic Review* 64(6): 950-963.

Lilleør, H.B. (2008a), "Can Future Uncertainty Keep Children out of School?", Working paper no 2008-06, Centre for Applied Microeconomtrics (CAM), University of Copenhagen.

Lilleør, H.B. (2008b), "Sibling Dependence, Uncertainty and Education. Findings from Tanzania", Working paper no 2008-05, Centre for Applied Microeconomtrics (CAM), University of Copenhagen.

Lilleør, H.B. (2008c), "Human Capital Diversification within the Household: Findings from Rural Tanzania", Working paper no 2008-04, Centre for Applied Microeconomtrics (CAM), University of Copenhagen.

Ministry of Education (2005), Education Sector Development Program III (ESDP-III): Program Action Plan, Addis Ababa.

Nerman, M., T. Owens (2010), "Determinants of Demand for Education in Tanzania: Costs, Returns and Preferences", Working Papers in Economics No. 472, Department of Economics, University of Gothenburg.

Orazem, P.F, E.M. King (2007), "Schooling in Developing Countries: The Roles of Supply, Demand and Government Interventions", chapter 55 in T. Schultz and J. Strauss (eds.) Handbook of development economics, volume 4, North Holland.

Oumer, J. (2009), The Challenges of Free Primary Education in Ethiopia", International Institute for Educational Planning (IIEP), Paris.

Pouliot W (2006). "Introducing Uncertainty into Baland and Robinson's Model of Child Labor", *Journal of Development Economics*, 79(1): 264-272.

Rosenzweig, M., Wolpin, K. (1985), "Specific Experience, Household Structure and Intergenerational Transfers: Farm Family Land and Labor Arrangements in Developing Countries", *Quarterly Journal of Economics* 100(Supplement): 961-987.

Schultz, T. (1975), "The Value of the Ability to Deal with Disequilibria", *Journal of Economic Literature* 13(3): 827-856.

World Bank (2005), "Education in Ethiopia: Strengthening the Foundation for Sustainable Progress", Washinton DC.

Appendix

Table A1: Variables and summary statistics

Variable	Obs	Mean	Std. Dev.
Age in years	11723	11.063	3.022
Year=2000	11723	0.142	0.349
Year=2001	11723	0.145	0.352
Year=2002	11723	0.149	0.356
Year=2003	11723	0.148	0.355
Year=2004	11723	0.147	0.354
Year=2005	11723	0.142	0.349
Year=2006	11723	0.128	0.334
Health shock	11695	0.177	0.381
Environmental shock	11695	0.346	0.476
Number of older sisters	11723	1.713	1.298
Number of older brothers	11723	1.858	1.379
Older sisters' literacy rate	9620	0.480	0.462
Older brothers' literacy rate	11690	0.513	0.430
Household head belonged to the least risk-averse category	11723	0.341	0.474
Household head belonged to the middle risk-averse category	11723	0.490	0.500
Household head belonged to the most risk-averse category	11723	0.169	0.375

Paper V

The Effect of Older Siblings' Literacy on School-Entry and Primary School Progress in the Ethiopian Highlands

Annika Lindskog

Department of Economics

School of Business, Economics and Law

University of Gothenburg

The effects of older sisters' and brothers' literacy on the annual school entry and primary school grade progress probabilities of boys and girls are estimated using within-household variation. Older siblings' literacy has positive effects, especially for same-sex siblings. The literacy of older sisters appears to be more beneficial than that of older brothers, not least since it has positive effects on school entry among both boys and girls, and since it has positive effects also when the sister has left the household. There are positive effects both from literate older siblings who left school and from literate older siblings who are still in school. This suggests that within-household education spillovers, rather than time-varying credit constraints, explain the positive sibling-dependency, since with credit constraints children in school would compete over scarce resources. The positive effects on school progress are limited to same-sex siblings who are still present in the household, suggesting every-day interactions to be important.

Keywords: Primary education, Ethiopia, Within-Household, Spillovers, Credit-Constraints

JEL-codes: I21, D13

1. Introduction

There is a general consensus on the importance of parents' education for the education of their children. Increasing the schooling of today's children should thus have positive spillovers on the schooling of future generations. But is there a more immediate effect of educating a child on the education of her siblings? Little is known about this, and the answer has implications for all policies directed towards the education of individual children.

The effect of older sisters' and brothers' literacy on girls' and boys' primary schooling is here investigated, using data from the rural Amhara region in Ethiopia, a place where until recently most individuals have had very limited experience with formal education.

The annual conditional school entry and school progress probabilities for 6-16 year old girls and boys from 2000 to 2006 are estimated. To control for unobserved parental characteristics, a linear probability model with household fixed effects is used. This is important since older siblings' education is clearly endogenous to parents' attitudes to education and child human capital investments.

Theoretically there are reasons to expect both positive and negative effects of siblings' education, making the direction of a possible effect an empirical question. First, the total effects of older siblings' literacy on the school entry and school progress probabilities are estimated. Next, an attempt is made to answer which mechanisms created the effects, focusing on the role of time-varying credit constraints and within-household spillovers. The focus is on these two mechanisms since the total effects turn out to be positive and they could create positive sibling-dependency.

To differentiate between the mechanisms, literate older sisters and brothers are divided into those who were still in school and those who had left school. With time-varying credit constraints, we would expect positive effects of older siblings who had left school, but negative effects of older siblings who were still in school. Positive within-household spillovers would be expected both if older siblings were in school and if they had left school. To evaluate the importance of everyday interactions, literate older siblings are also divided into those who were still living in the household and those who were not.

Previous studies have investigated the effect of child-specific conditional cash transfers on education of siblings, where the cash transfer programme is an external intervention that should be unrelated to unobserved parental characteristics. In spite of an expected positive income effect, Barrera-Osorio et al. (2008), using data from Colombia, find a negative effect on schooling of having a sibling in the programme. Possible reasons could be a reallocation of household responsibilities across siblings or a diversification motive. Ferreira et al. (2009) find no effects of a Cambodian cash-transfer programme on schooling of siblings, even if the cash transfer was of a relatively smaller size than in Colombia. Both programmes were targeted towards increased secondary education.

While the literature on schooling effects of siblings' education is sparse, that on the effects of number and gender of siblings, and of birth order in the sibship, is large, despite the fact that number of siblings, just as education, is likely to be endogenous to unobserved parental preferences. According to the hypothesis on a child quantity-quality tradeoff, parents will beforehand make decisions about the number of children and how much to invest in their human capital (Becker and Lewis, 1973).

Sibling gender and relative birth order should be largely exogenous.[1] In developing countries, it is often found to be advantageous to be a later-born in the sibship. This is consistent with an impeding role of poverty and credit constraints for children's schooling, since parents' income tends to increase over time (Parish and Willis, 1993) and since older siblings can contribute to household income when they start working (Emerson and Souza, 2002; Chesnekova and Vaithianathan, 2008). Some authors have found a positive effect of having older sisters rather than brothers, and have therefore suggested that older sisters' work often makes younger siblings' education possible (Parish and Willis, 1993; Glick and Sahn, 2000; Morduch, 2000).

In developed countries, a negative effect of number of siblings, consistent with a child quantity-quality tradeoff, is often found. Also consistent with a quantity-quality tradeoff is the fact that the negative effect is usually reduced when instruments such as twin-births are used (Black et al., 2005; Booth and Joo Kee, 2009). In developing countries, and especially in Africa, it is, however, common to find positive sibship

[1] Sibling gender might not be completely exogenous in countries with strong son preferences and gender-selective abortions. However, this is not relevant in Africa.

effects (Chernichovsky, 1985). Some argue that this could be explained by the important role of the extended family in Africa; the additional costs of many children are absorbed in the larger extended family. Cornwell et al. (2005) find that more school-age children increases schooling for black South-African teenagers, and explain this with economies of scale in schooling and the creation of a 'culture of schooling'.

The remainder of this study starts with a theoretical discussion on mechanisms that could create sibling-dependency in education. This is followed by a section on the data, and one on the empirical approach. Section 5 presents the empirical results and Section 6 a discussion, and Section 7 concludes the paper.

2. What could create sibling-dependency in education?

Economic theory on education is dominated by human capital theory, which stipulates that investments in children's schooling are made as long as the marginal benefit exceeds the marginal cost (Becker, 1962; Ben-Porath, 1967). In the simple model, neither the presence of siblings nor their education matter for schooling.

Market imperfections, however, change this simple picture. Credit constraints have received most attention in the literature (Jacoby, 1994; Edmonds, 2006a; Gitter and Barham, 2007). Without perfect access to credit the schooling decision cannot be separated from decisions on the inter-temporal allocation of consumption; a poor household might not be able to afford costly schooling investments in spite of high expected future returns. With credit constraints, competition for scarce resources in the household creates negative sibling-dependency in education, especially among closely spaced siblings (Jacoby, 1994; Morduch, 2000)[2]. On the other hand, credit constraints can be eased by higher income contributions from better educated older siblings (Emerson and Souza, 2008; Chesnekova and Vaithianathan, 2008). With credit constraints, there are thus reasons to expect both positive and negative effects of older siblings' education, depending on whether or not they are still in school.

[2] There is also a large literature on competition over scarce resources, or resource dilution, in sociology where authors tend to have both limited financial resources and parental time in mind (Blake, 1989).

Labour market imperfections are also of relevance. Household labour is likely to be either the only source of household and farm labour available or at least considerably more cost-effective than hired-in labour. In a seminal paper, Bhalotra and Heady (2003) show how labour market imperfections could make child work more frequent in households owning more land, since the marginal product of (child) labour is higher in these households. Along the same lines, Edmonds (2006b) argues that education inequality across siblings could often be explained by differences in marginal productivity of child time at alternative activities. If household and farm work is shared by household members, the schooling decisions for different siblings cannot be separated from each other and neither can they be separated from decisions on household and farm production. It has for example been found that the presence of pre-school age children decreases schooling for girls, while the presence of other school-age girls has the opposite effect. These links are most likely due to the different effects on girls' household responsibilities (Glick and Sahn, 2000). If work is shared by household members and the time spent in school is increased for one child, the consequence could be that the burden of household duties and farm work is increased for other children in the household (Barrera-Osorio et al., 2008). While the presence of older siblings in the household should be beneficial, given their presence, their schooling coulde have a negative effect on the education of younger siblings.

Since the presence of siblings and their education should not matter in the simple human capital model, many have argued that any presence of sibling effects indicates credit constraints (Jacoby, 1994; Morduch, 2000; Sawada and Lokshin, 2009). Yet, as described, sibling effects could also be the outcome of differences in marginal productivity of time when labour markets are imperfect. However, another possibility is that siblings and their education directly affect costs and benefit, either actual or perceived or both. Such effects will hereinafter be referred to as within-household spillovers.

One source of such spillovers could be economies of scale in schooling; the additional cost of having one child more in school could be lower than the cost of having the first child in school (Cornwell et al., 2005). Children could accompany each other to school, and they may share books, clothes or school uniforms. Children could also enhance each other's learning and thus increase the benefits of schooling.

Using examples from Bangladeshi and Anglo children living in East London, Gregory (2001) finds evidence of sibling 'synergies', where older children teach younger siblings and by doing so also develop their own learning.

Spillovers could also be the result of siblings' influence on each other's and parents' attitudes and beliefs. The perceptions of costs and benefits of education could change in view of the siblings' experience of education. Beenstock et al. (2008) argue that interaction between siblings, which could supposedly both enhance learning and influence beliefs and attitudes, is an important factor behind sibling correlation in schooling and in earnings in Israel. Cornwell et al. (2005) argue that a 'culture of schooling' can be created in the household when one child is educated. They also suggest that a positive effect of migrant household members on girls' schooling could be explained by the increased knowledge about the benefits of schooling. In an experiment in a neighbourhood with few well-educated individuals in Santo Domingo, the Dominican Republic, Jensen (2010) shows the importance of perceived benefits of education for schooling decisions and how provision of information on returns to education can increase schooling. The effects of siblings' education on perceived benefits might be especially important in a setting such as rural Amhara, with very limited experience of formal education.

The studies on the effect of child-specific conditional cash transfers on the education of siblings in Colombia (Barrera-Osorio et al., 2008) and Cambodia (Ferreira et al., 2008) were mentioned earlier. One reason for the more beneficial outcome of the Cambodian programme could be that the money was received during ceremonies where the benefits of formal education were stressed. This might well have changed parents' view on the value of schooling.

Lastly, the returns to education are uncertain, and the riskiness associated with human capital investments could mean that parents want to diversify. This would result in negative sibling-dependency (Lilleør, 2008). In the fourth paper of this thesis, a possible desire to diversify household human capital investment is specifically investigated, by studying sibling-dependency in education in households with differently risk averse head, using the same Ethiopian data as in the present paper.

3. The data on children's primary schooling in rural Amhara

The data used in this study is from the Ethiopian Environmental Household Survey (EEHS). The sampled households come from 13 Kebeles[3] in the South Wollo and East Gojjam zones of the Regional State of Amhara. The two zones were chosen to represent different agro-climatic zones in the Ethiopian highlands, with less rainfall in South Wollo than in East Gojjam. Most households, in the sample and in the study areas, make their living on rain-fed subsistence agriculture. Access to roads and capital markets is limited in the areas.

To date, four rounds of data have been collected, in 2000, 2002, 2005 and 2007. Most of the information on children's education was collected in the fourth round, where respondents were asked about the schooling history of all household members age 6 to 24. Data was also collected for household members no longer residing in the household, yet less successfully.

The fourth round of data was used to create an annual panel, from 2000 to 2006, on entry into first grade and primary school progress. Interviews were conducted in April/May, i.e. at the end of the Ethiopian school year (September-June). Since children enter school in September and complete grades in June, there is information on school entry and grade progress only until 2006. The panel starts in 2000 since the first round of data collection took place in April/May that year, and explanatory variables use information from the last preceding round.

In the total sample from the fourth round, there are 5549 children from 1674 households. Disregarding children without older siblings reduces the sample to 2783 children from 936 households. Leaving out first-born girls and boys clearly means that the sample is not representative of all children in rural Ethiopia, but only of children with older sisters and brothers. Moreover, households with young and small families are more likely to be excluded from the sample. If small families are different in terms of human capital investments, as suggested by the quantity-quality tradeoff hypothesis, the sample becomes unrepresentative also at the household level. A comparison of households excluded from the sample due to lack of children with

[3] A Kebele is the smallest administrative unit in Ethiopia. In rural areas it is more or less equivalent to a village, while in larger cities it comprises a neighbourhood.

older brothers and sisters with other households (in Table 1) shows that the excluded households are both smaller and younger, but they do not differ with respect to literacy of the household head and spouse or school entry and progress of young household members.[4]

Table 1: Comparison of households without children with older sisters and brothers with other households

	Excluded households		Included Households			
	Mean	St. dev.	Mean	St. dev.	Difference	St. dev.
Entry rate	0.208	0.183	0.193	0.148	0.016	0.011
Progress rate	0.931	0.142	0.936	0.123	-0.005	0.010
Literate head	0.429	0.495	0.447	0.497	-0.019	0.032
Literate spouse	0.214	0.406	0.215	0.411	-0.000	0.028
Household size	3.057	2.489	6.079	1.907	-3.021***	0.142
Age of head	46.711	16.587	49.573	12.035	-2.862***	0.993
Asset index	12.705	28.003	15.190	27.559	-2.484	1.802

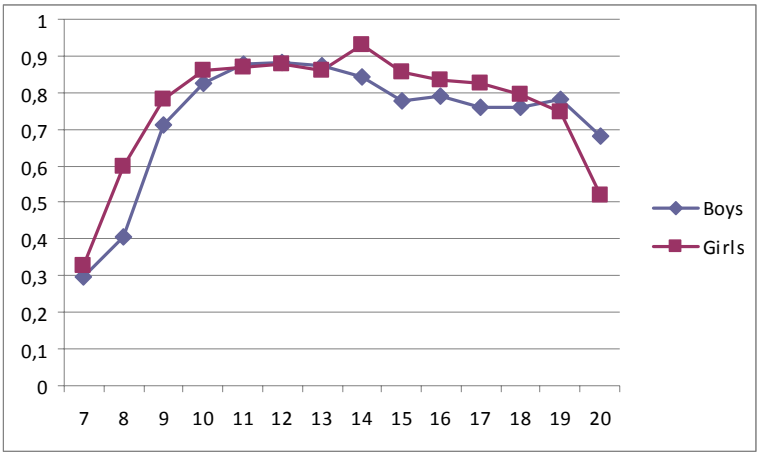
* = p<0.10, **= p<0.05, ***= p<0.01

There are 2767 children with non-missing school entry data from 2000 to 2007, and 1173 children with non-missing school progress data. Of the children without progress data, 36.2% had never started school, data is inconsistent for 34.3% and data is missing for 29.5%.

Figure 1 shows the share of boys and girls of different ages who had ever attended school in 2007. Late entry is common among both girls and boys, but more so among boys. It is also clear that schooling increased relatively recently: after age 13 for boys and age 15 for girls, the share who had attended decrease with age.

[4] A similar comparison was made in the fourth thesis paper, with slightly different figures. Figures differ both since the fourth paper comparison use only a sample of households whose heads had relatively stable risk preferences, and since it compares households where no one had an older sibling with other households, while households where no one had both an older sister and an older brother is compared with other households here.

Figure 1: Share of boys and girls who had ever attended school in 2007, by age.



The increasing share of children who attended school is a result of a massive primary school expansion in Ethiopia since the mid-1990s; from 1994/95 to 2005/06 the primary school gross enrolment rate in Ethiopia rose from 34.0% to 91.3%, and the gender parity index narrowed from 0.6 in 1997/98 to 0.84 in 2005/2006. In Amhara the net enrolment in 2004/2005 was 54.6% for boys and 53.1% for girls, making Amhara one of few regions with similar enrolment rates for girls and boys. As is common with such large expansions in enrolment though, the numbers of teachers and classrooms have not increased at pace with the number of pupils, raising concerns about reduced quality (Oumer, 2009; Ministry of Education, 2005; World Bank, 2005).

4. Empirical model

Completed schooling is the outcome of several sequential decisions: first on school entry then on school continuation. In this paper we separately estimate the annual conditional school entry and primary school grade progress probabilities.

The official school start age in Ethiopia is 7, but some start at age 6, and it is also common to start later, particularly in rural areas. In the empirical analysis, a child is classified as eligible to enter school if he or she is 6-16 years old and has not started previously. If someone who is eligible to enter school does not, it means that he or she

will never start school or will start later. A child is eligible for primary school grade progress if he or she is in primary school and is aged 6-16 years old. Failure to progress is due either to grade repetition or due to quitting school.

Essentially, the annual conditional school entry and grade progress probabilities are discrete-time "hazards". The advantage of estimating the annual school entry probability rather than duration until (or age at) school entry is that the data can be fully used while avoiding censoring problems. There is no need to restrict the sample to children old enough to know that most of those who haven't entered will not do so later. This is especially useful for the results to remain relevant in a situation, such as Ethiopia, where schooling has increased massively in recent years.

The main explanatory variable, literacy of older siblings, is clearly endogenous to parental characteristics that affect decisions on education of all children in a family. Some of these characteristics, such as parental attitudes towards formal education and investment in child human capital in general (education and health), are unobservable. To control for unobserved household characteristics, I use household fixed effects in a linear probability model. The fixed effects will capture the effect of many variables commonly considered to be important for children's schooling, e.g. permanent income and education of the household head and spouse.

There may still be problems of time-varying shocks to the households, affecting the education of all children in a household. Older siblings' education is lagged, something that would deal with shocks in a model without fixed effects. Yet with fixed effects, there must be strict exogeneity; that is, the explanatory variables should be uncorrelated with lags and leads of the error term (Arellano, 2003). In an attempt to deal with time-varying shocks, I include health and environmental shocks among the explanatory variables. Health shocks are deaths or illnesses of household members, and environmental shocks are mainly floods and droughts, but also other weather-related shocks and pests that hit plants or animals. The shock dummies measure whether a shock occurred at least once during the period between survey rounds.[5]

[5] If there was a shock during 2005-2006, the dummy was set to equal one for both these years. While there is information about the timing of the last shock in the data, there is no information about the timing of earlier shocks, so an annual shocks series could not be created. Estimations using an index of wealth in the preceding round instead of shocks where also run, which did not qualitative affect results.

Older siblings' education is measured by their literacy. Though a rough indicator of education, it has the advantage of few missing values. Literacy is measured by the number of literate older siblings. Since also the total number of older siblings is included this measures their literacy. The effect of older siblings could depend on gender, hence, older sisters are differentiated from older brothers and estimations are done separately for girls and boys. The number of older sisters and brothers and the number of older literate sisters and brothers are from the last preceding survey round and include both siblings who were living in the household and siblings who had left. In some estimations, a distinction is made between older literate siblings who were still in school and those who had left school and between older literate siblings who were living in the household and those who had left the household.

The model also controls for the year and for child age in years. Year dummies capture both the effect of the massive expansion of primary education during the study period and the possible effect of aging of households in the panel. Table A1 in the appendix report summary statistics of all variables.

5. Results

This section first presents results from the main regressions of the annual school entry and school progress probabilities on literacy of the child's older sisters and brothers (Table 2). Thereafter Table 3 reports regressions differentiating between literate siblings who were still in school and those who had another main activity, and Table 4 reports regressions differentiating between literate siblings who were still living in the household and those who had left.

Literacy of older siblings positively influences schooling of younger siblings. Both girls' and boys' school entry probabilities are more positively affected by older sisters' literacy than by older brothers' literacy; if an older sister is literate the school entry probability increases by 0.082 for girls and by 0.090 for boys, while if an older brother is literate the school entry probability increases by 0.041 for girls and by a smaller and not statistically significant amount for boys.

Since income and wealth could be endogenous to older siblings' education, the shocks variables were preferred despite their limitations.

Table 2: Effect of older siblings' literacy on primary schooling – Coefficients from linear probability models with household fixed effects.

	The annual school entry probability		The annual primary school progress probability	
	Girls	Boys	Girls	Boys
	(2)	(1)	(4)	(3)
Age in years	0.018***	0.033***	-0.016**	-0.012*
	(0.007)	(0.006)	(0.008)	(0.007)
Year 2001	0.074**	0.069**	0.020	0.001
	(0.029)	(0.027)	(0.029)	(0.028)
Year 2002	0.110***	0.116***	0.004	0.026
	(0.035)	(0.031)	(0.034)	(0.032)
Year 2003	0.206***	0.167***	-0.032	0.033
	(0.039)	(0.035)	(0.038)	(0.031)
Year 2004	0.174***	0.126***	-0.024	0.059
	(0.049)	(0.045)	(0.044)	(0.039)
Year 2005	0.250***	0.230***	-0.018	0.056
	(0.051)	(0.044)	(0.049)	(0.042)
Year 2006	0.281***	0.169***	-0.001	0.074
	(0.059)	(0.050)	(0.053)	(0.046)
Health shock	0.006	0.006	-0.008	0.011
	(0.041)	(0.039)	(0.026)	(0.017)
Environmental shock	-0.037	0.007	-0.018	0.007
	(0.035)	(0.036)	(0.020)	(0.016)
Number of older sisters	0.002	-0.004	0.010	-0.007
	(0.022)	(0.020)	(0.017)	(0.013)
Number of older brothers	-0.014	0.055***	-0.013	-0.027
	(0.028)	(0.020)	(0.028)	(0.018)
Literate older sisters	0.082***	0.090***	0.050**	0.019
	(0.031)	(0.028)	(0.022)	(0.024)
Literate older brothers	0.041*	0.022	-0.004	0.040**
	(0.025)	(0.021)	(0.020)	(0.016)
Constant	-0.151	-0.444***	1.090***	1.048***
	(0.102)	(0.090)	(0.101)	(0.090)
Observations	2114	2382	1865	1910
Children	693	785	515	564
Housholds	496	540	398	425

Standard errors, clustered at the household, in parentheses.
All estimations also include age, year dummies, a household wealth index, and a constant.
* = p<0.10, **= p<0.05, ***= p<0.01

Among girls, the primary school progress probability is more positively influenced by older sisters' than older brothers' literacy (0.050 compared to a close to zero and statistically insignificant effect), while among boys, it is more positively

influenced by older brothers' literacy (0.040 compared to a smaller and statistically insignificant effect).[6]

The number of older brothers has a positive effect on boys' school entry. But since first-born boys are not included in the sample, this says nothing about the difference between being first born and later born. Also, since only within-household variation is used for identification, it says nothing about the difference between families with more or fewer children. Among a given number of brothers, it thus appears beneficial to be as young as possible.

Turning to other explanatory variables, age increases the annual school entry probability, reflecting that late school entry is common. And, as should be expected, age decreases the annual school progress probability, since older children are more likely to quit school. The omitted reference year is 2000, and the coefficients of year dummies thus show that the probability that a child enters school increased during 2000-2006, and more so for girls than for boys. However, school progress does not seem to have changed much during 2000-2006. Last, there are no statistically significant effects of the shock variables.

Older siblings' literacy seems to have beneficial effects on younger siblings' primary schooling for the children in rural Amhara. But why is this so? In the theoretical discussion, two mechanisms that could cause positive sibling dependency were brought up; time-varying credit constraints and within-household spillovers.

With credit constraints we would expect older siblings still in school to have a negative effect on primary schooling of younger siblings. But we would expect a positive effect of older siblings who have left school and who have started to contribute to household income. Positive within-household spillovers, on the other hand, could arise both from older siblings in school and from literate older siblings who have left school. In the regressions presented in Table 3, a distinction is therefore

[6] For boys, the null hypothesis of an equal effect of older sisters' and older brothers' literacy is rejected at the ten percent level for school entry (F-statistic=3.73), while it cannot be rejected for school progress. For girls, the null hypothesis is instead rejected at the ten percent level for school progress (F-statistic=2.94), while it cannot be rejected for school entry.

made between literate siblings who were still in school and those with another main activity.[7]

Table 3: Effect of literate siblings who were still in school and of those who had left school on primary schooling: Coefficients from linear estimations with household fixed effects.

	The annual school entry probability		The annual school progress probability	
	Girls	Boys	Girls	Boys
	(2)	(1)	(4)	(3)
Number of older sisters	0.005	-0.007	0.010	-0.006
	(0.022)	(0.020)	(0.017)	(0.013)
Number of older brothers	-0.008	0.052**	-0.013	-0.025
	(0.030)	(0.021)	(0.028)	(0.017)
Literate sisters in school	0.089**	0.075**	0.030	0.038
	(0.038)	(0.036)	(0.030)	(0.029)
Literate brothers in school	0.026	0.030	-0.006	0.035**
	(0.030)	(0.025)	(0.024)	(0.016)
Literate sisters who had left school	0.083***	0.090***	0.051**	0.021
	(0.031)	(0.028)	(0.021)	(0.024)
Literate brothers who had left school	0.038	0.023	-0.004	0.040**
	(0.025)	(0.021)	(0.020)	(0.016)
Observations	2114	2382	1865	1910
Children	762	853	524	587
Households	532	564	404	437

Standard errors, clustered at the household, in parentheses.
All estimations also include age, year dummies, a household wealth index and a constant.
* = p<0.10, **= p<0.05, ***= p<0.01

Whether or not an older sibling is still in school, the effects of their literacy appear to be positive – not negative. Moreover, the beneficial effects are fairly similar for older siblings who were in school and literate older siblings who had left school. Hence, positive sibling-dependency cannot be explained only with time-varying credit constraints – spillovers are also of importance. If an older sister was in school rather than illiterate, the annual school entry probability increases by 0.089 for girls and 0.075 for boys. If she was literate and had left school, the annual school entry probability increases by 0.083 for girls and 0.090 for boys. However, there is also a statistically significant increase in girls' annual school progress probability when the literate older sister had left school. If an older brother was in school rather than illiterate, boys' school progress probability increases by 0.035. If a literate older brother had left school, boys' school progress probability increases by 0.040.

[7] In the fourth round the most common activities of non-household heads, age 15-29, who did not study, were: farming (41%), domestic work (28%), non-farm work (17%), and looking for work (10%).

The regressions presented in Table 4 differentiate between literate older siblings who were still living in the household and those who had left. How do the effects depend on the presence of a sibling in the household? There should be a larger potential for positive interactions between siblings if they live together. If such everyday interactions are important for beneficial effects of older siblings' literacy on children's primary schooling, the beneficial impacts should of course mainly come from siblings in the household. However, we cannot with certainty differentiate between the importance of time-varying credit constraints and positive within-household spillovers with the results in Table 4. Larger effects of siblings in the household could be due to higher income contributions when they are still part of the household. Larger effects of siblings who have left the household could be due to changed perceptions of the benefits of schooling if the literate older sibling seems to be doing well, or it could be due to income contributions to the household of origin.

Table 4: Effects of literate siblings who lived in the households and of those who had left on primary schooling: Coefficents from linear estimations with household fixed effects.

	The annual school entry probability		The annual school progress probability	
	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)
Number of older sisters	-0.004	0.001	-0.007	0.010
	(0.020)	(0.022)	(0.013)	(0.017)
Number of older brothers	0.056***	-0.013	-0.024	-0.015
	(0.020)	(0.027)	(0.018)	(0.029)
Literate sisters in the household	0.084***	0.079**	0.031	0.051**
	(0.030)	(0.032)	(0.029)	(0.022)
Literate brothers in the household	0.026	0.056**	0.041**	-0.008
	(0.022)	(0.026)	(0.016)	(0.019)
Literate sisters that left the hh	0.193	0.487***	-0.104	0.049
	(0.132)	(0.154)	(0.099)	(0.033)
Literate brothers that left the hh	-0.021	-0.103	0.017	0.033
	(0.051)	(0.066)	(0.055)	(0.073)
Observations	2382	2114	1910	1865
Children	853	762	587	524
Households	564	532	437	404

Standard errors, clustered at the household, in parentheses.
All estimations also include age, year dummies, a household wealth index and a constant.
* = p<0.10, **= p<0.05, ***= p<0.01

Literate older sisters who lived in the household positively influenced school entry of both girls and boys, but only school progress of girls. If an older sister was literate and living in the household, the annual school entry probability increased by

0.079 for girls and by 0.084 for boys. Girls' annual school progress probability increased by 0.051. Literate older brothers who lived in the household had a positive effect on school entry for girls, increasing the annual school entry probability by 0.056, and on school progress for boys, increasing the annual school progress probability by 0.041.

Literate older sisters who had left the household has a huge positive effect on girls' school entry, increasing the annual probability of school entry by 0.487. There was also a large estimated effect on boys' school entry (0.193), yet it was not statistically significant. There were no statistically significant effects of literate older brothers who had left the household. Also, there were no positive effects of literate older siblings, regardless of gender, who had left the household; only of same-sex siblings still living in the household.

6. Discussion

First on annual school entry probabilities, literacy of older sisters seems to be more beneficial than literacy of older brothers for both boys and girls. Moreover, literate older sisters, as opposed to brothers, have beneficial effects even if they have left the household. Although there is a positive effect of older sisters in the household on girls' school entry, the effect of sisters who have left the household is much larger. Also for boys, the estimated effect of sisters who have left the household is larger than that of sisters still living in the household, yet the coefficient has a large standard error and is not statistically significant. The beneficial effect of older sisters who had left suggests that older sisters fare better when leaving the household if they are literate, something that could both increase their possibility to help the household of origin (easing credit constraints) and serve as a positive example of the benefits of schooling, in particular for girls.

There are different possible reasons as to why literate older sisters could fare well when leaving the household of origin. First, their own income earning capacity probably increases. The fact that over 50 percent of female household members who left the household age 10-25 did so to marry suggests that it is also important how literate girls manage in the marriage market. Fafchamps and Quisimbing (2005) find

substantive assortative matching in the marriage market in rural Ethiopia, where matching on human capital of spouses is becoming increasingly important. They also show that the lion's share of resources of the newly formed household comes from the groom and his family, and that gifts from the groom's to the bride's family are comparatively small. Hence, better educated older sisters probably end up in a wealthier new household. Moreover, their bargaining power in the new household is probably larger. Fafchamps et al. (2009) suggest cognitive ability to be important for intra-household bargaining power in rural Ethiopia.

Literate older brothers positively influence the annual school entry probability of girls, but only if the older brothers still live in the household. When differentiating between literate older brothers in school and literate older brothers who have left school, both coefficients become statistically insignificant. Since presence of an older brother in the household matters and the effect does not appear to differ depending on whether they are still in school, some sort of spillover seems to be a more likely explanation than income contributions easing credit constraints.

The number of older brothers has a positive effect on boys' school entry. To have more older brothers is good for boys, and it does not matter if they are literate or not. One possible explanation for this is that sons are expected to work on the family farm and later inherit it. With more sons, the work of the youngest son might not be needed, and parents might prefer to give him more formal education instead of preparing him to inherit the farm, i.e. behind this could be a low current opportunity cost of the boys' time, a low expected return to learning by doing in the field (if he is not expected to inherit the land), and a household-level diversification motive.

Turning to annual school progress probabilities, the positive effects of older siblings' literacy were limited to same-sex siblings, and while they did not differ depending on whether the older sibling was still in school or not, they do seem to depend on the presence of the older sibling in the household. This suggests that older siblings' income contributions are not important in this context, but that everyday interactions are. Interaction with same-sex older literate siblings can probably enhance a child's learning. Such effects on learning should be more important for school progress than for school entry. Everyday interactions could also have important effects on attitudes and beliefs.

The positive effects of older siblings' literacy on younger siblings' schooling found in this study contradict the results from studies on the impact of child-specific conditional cash transfers on siblings' education in Colombia (Barrera-Osorio et al., 2008) and Cambodia (Ferreira et al., 2008). These studies of course differ from the present one in many important ways, but a potentially central one that may explain the difference in results is that they estimate immediate effects while this study is likely to capture more long-run effects. Reallocation of household work duties is for example likely to matter mostly in the short run. Another potentially important difference is the very limited experience with any level of formal education in rural Amhara, and that the above studies investigate the effect of increased secondary education while the present study investigates the effect of a very modest level of education (literacy). The effect of a modest level of education in a place with little formal education may be very different from that of more secondary education in places where it is common to have at least some primary education.

7. Summary and conclusions

This paper investigates the effects of older sisters' and brothers' literacy on younger siblings' annual school entry and primary school progress probabilities, using only within-household variation and distinguishing girls from boys. Theoretically, there are reasons to expect both positive and negative effects.

Older siblings' literacy turns out to have positive effects on primary schooling of younger siblings. Overall, the beneficial impact of an older sister's education appears to be larger than that of an older brother's education. If an older sister is literate, the annual school entry probability increases by 0.082 for girls and by 0.090 for boys. If an older brother is literate, girls' annual school entry probability increases by 0.041. Boys' annual school entry probability instead increases when they have more older brothers, independent of their literacy. This could be since the boys' work is not needed on the farm and since they are not, as boys with fewer older brothers, expected to inherit the land.

Beneficial effects on school progress are limited to same-sex siblings, i.e. older sisters' literacy favours girls' school progress, increasing the annual probability by

0.050, and older brothers literacy favours boys' school progress, increasing the annual probability by 0.040.

So what can explain the positive effects of older siblings' literacy? From the theoretical discussion, two candidate explanations are time-varying credit constraints, eased by higher income contributions from better educated older siblings, and within-household spillovers, affecting actual or perceived benefits and costs of schooling (siblings could for example share books, walk or ride to school together, enhance each other's learning, and affect each other's beliefs about the benefits of schooling). With the credit-constraint explanation, we expect a negative effect of older literate siblings who are still in school (due to competition over scarce resources), but a positive effect if they have left school. A distinction is therefore made between literate siblings who are still in school and those who are mainly involved in something else. Moreover, to evaluate the importance of everyday interactions, regressions differentiating between literate older siblings who live in the household and those who have left it are made.

The effects of literate older siblings in school and of those who have left school turn out to be similar, suggesting an important role of spillovers. Moreover, effects of same-sex older literate siblings on school progress appear to depend on their presence in the household, suggesting an important role of everyday interaction. Girls' interaction with older literate sisters and boys' interaction with older literate brothers could probably enhance their learning, which should be more important for school progress than for school entry.

While literacy of older brothers seems to have beneficial effects on schooling only if they still live in the household, literacy of older sisters appears to matter also if they have left the household. Literate older sisters who have left have a huge positive effect on girls' annual school entry probability, increasing it by 0.487; the estimated effect on boys' school entry is also large (0.193), but statistically insignificant. This suggests that literate older sisters fare better than illiterate ones after leaving the household, making it possible for them to help their household of origin, but possibly also serving as a good example of the benefits of schooling, especially for girls.

Though there is little evidence on within-household spillovers other than from parents in the literature, it is not surprising to find them considering the large effects

of parents' education usually found. However, it is quite possible that interaction with educated older siblings has a larger effect when parents have little education, i.e. that parents' and older siblings' education are substitutes. Under circumstances similar to those in rural Ethiopia in the last 15 years, where education has been heavily expanded from a very low initial level, older siblings' education may be especially important. Thus, there are reasons to do research on within-household spillovers under different circumstances.

8. References

Barrera-Osorio, F., M. Bertrand, L. Linden, F. Perez-Calle (2008). "Conditional Cash Transfers in Education Design Features, Peer and Sibling Effects Evidence from a Randomized Experiment in Colombia", NBER Working Paper No. 13890.

Becker, G. (1962), "Investment in Human Capital: A Theoretical Analysis", *Journal of Political Economy* 70(5): 9-49.

Becker, G., H. Lewis (1973), "On the Interaction between the Quantity and Quality of Children", *Journal of Political Economy* 81(2): S279-S288.

Beenstock, M. (2008), "Deconstructing the Sibling Correlation: How Families Increase Inequality", *Journal of Family Economic Issues* 29: 325-345.

Ben-Porath, Y. (1967), "The Production of Human Capital and the Life Cycle of Earnings", *The Journal of Political Economy* 75(4): 352-365.

Bhalotra, S., C. Heady (2003)."Child Farm Labor: The Wealth Paradox", *World Bank Economic Review* 17 (2): 197-227.

Black, S., P. Devereux, K. Salvanes (2005), "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education", *Quarterly Journal of Economics* 120(2): 669-700.

Blake, J. (1989), "Family Size and Achievement", University of California Press, Los Angeles.

Booth, A., H. Kee (2009), "Birth Order Matters: the Effect of Family Size and Birth Order on Educational Attainment", *Journal of Population Economics* 22(2): 367-397.

Chernichovsky, D. (1985), "Socioeconomic and Demographic Aspects of School Enrollment and Attendence in Rural Botswana", *Economic Development and Cultural Change* 33(2): 319-332.

Chesnokova, T., and R. Vaithianathan (2008), "Lucky Last? Intra-Sibling Allocation of Child Labor", *The B.E. Journal of Economic Analysis & Policy* 8(1).

Cornwell, K., B. Inder, P. Maitra, A. Rammohan (2005),. "Household Composition and Schooling of Rural South African Children: Sibling Synergy and Migrant Effects", mimeo.

Edmonds, E (2006a), "Child Labor and Schooling Responses to Anticipated Income in South Africa", *Journal of Development Economics* 81(2): 386-414.

Edmonds, E (2006b), "Understanding Sibling Differences in Child Labor", *Journal of Population Economics* 19(4):795-821.

Fafchamps, M., A. Quisumbing (2005), "Assets at Marriage in Rural Ethiopia", *Journal of Development Economics* 77(1): 1-25.

Fafchamps, M., B. Kebede, A. Quisumbing (2009), Intrahousehold Welfare in Rural Ethiopia, *Oxford Bulletin of Economics and Statistics* 71(4), pp. 567–599.

Ferreira, F., D. Filmer, N. Schady (2009), "Own and Sibling Effects of Conditional Cash Transfer Programs: Theory and Evidence from Cambodia", World Bank Policy Research Working Paper No. 5001.

Gitter, S.R., B.L. Barham (2007), "Credit, Natural Disasters, Coffee, and Educational Attainment in Rural Honduras", *World Development* 35(3): 498-511.

Glick, P., D. Sahn (2000), "Schooling of Girls and Boys in a West African Country: the Effects of Parental Education, Income, and Household Structure", *Economics of Education Review* 19(1): 63-87.

Gregory, E. (2001), "Sisters and Brothers as Language and Literacy Teachers: Synergy between Siblings Playing and Working Together", *Journal of Early Childhood Literacy* 1 (3): 301-322.

Jacoby, H. (1994), "Borrowing Constraints and Progress through School: Evidence from Peru", *Review of Economics and Statistics* 176(1): 151-160.

Jensen, R. (2010), "The Perceived Returns to Education and Demand for Schooling", *The Quarterly Journal of Economics* 125(2): 515-548.

Lilleør, H.B. (2008), "Sibling Dependence, Uncertainty and Education. Findings from Tanzania", Working paper no 2008-05, Centre for Applied Microeconomtrics (CAM), University of Copenhagen.

Ministry of Education, (2005), "Education Sector Development Programme III (ESDP-III): Program Action Plan", Addis Ababa.

Morduch, J. (2000), "Sibling Rivalry in Africa", *American Economic Review* 90(2): 405-409.

Oumer, J. (2009), "The Challenges of Free Primary Education in Ethiopia", International Institute for Educational Planning (IIEP), Paris.

Parish, W., R. Willis (1993),"Daughters, Education, and Family Budgets: Taiwan Experiences", *Journal of Human Resources* 28(4): 863-898.

Sawada, Y., M. Lokshin (2009), "Obstacles to School Progression in Rural Pakistan: An Analysis of Gender and Sibling Rivalry using Field Survey Data", *Journal of Development Economics* 88(2): 335-347.

World Bank (2005), "Education in Ethiopia: strengthening the foundation for sustainable progress", World Bank, Washinton DC.

Appendix

Table A1: Variables and summary statistics

Variable	Obs	Mean	Std. Dev.
Age in years	8081	11.116	2.977
Year=2000	8081	0.117	0.322
Year=2001	8081	0138	0.345
Year=2002	8081	0.143	0.350
Year=2003	8081	0.141	0.348
Year=2004	8081	0.163	0.369
Year=2005	8081	0.156	0.363
Year=2006	8081	0.141	0.348
Health shock	8059	0.178	0.383
Environmental shock	8059	0.347	0.476
Number of older sisters	8081	2.138	1.180
Number of older brothers	8081	2.277	1.281
Number of literate older sisters	8081	0.778	0.891
Number of literate older brothers	8081	1.416	1.165
Number of literate older sisters who were in school			
Number of literate older brothers who were in school			
Number of literate older sisters who lived in the household			
Number of literate older brothers who lived in the household			

Paper VI

Journal of Economic Behavior & Organization 72 (2009) 884–902

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Preferences for redistribution—A country comparison of fairness judgements

Ann-Sofie Isaksson, Annika Lindskog*

Department of Economics, University of Gothenburg, Box 640, 405 30 Göteborg, Sweden

ARTICLE INFO

Article history:
Received 6 May 2008
Received in revised form 11 August 2009
Accepted 11 August 2009
Available online 19 August 2009

JEL classification:
D63
D31
D01

Keywords:
Redistribution preferences
Fairness

ABSTRACT

This paper seeks to explain within- and between-country variation in redistributive preferences in terms of self-interest concerns and an input-based concept of fairness, which we examine by looking at the effects of beliefs regarding the causes of income differences. Results of estimations based on data for 25 countries indicate that both factors are indeed important determinants of redistribution support, in line with hypothesised patterns. We find that while differences in beliefs on what causes income differences seem to be important for explaining within-country variation in redistributive preferences, they do little to explain between-country differences. Differences in the *effects* of holding certain beliefs, however, are important for explaining between-country variation in redistributive preferences, suggesting considerable heterogeneity across societies in what is considered as fair.

1. Introduction

Rational economic self-interest fails to explain the wide spread in support for income redistribution.[1] Judging from standard economic reasoning, according to which individuals are motivated by self-interested utility maximization, this is puzzling. However, based on a vast experimental literature there is a growing consensus that people are motivated by forces other than self-interest, and particularly so by fairness considerations.[2]

One could in this context make a distinction between fairness concepts focusing only on outcomes, such as strict egalitarianism,[3] and those accounting for individual inputs contributing to those outcomes. The general idea that the fair distribution should depend on individual inputs is well established, both in the normative literature on justice and in positive analyses of what people consider to be just. According to equity theory dating back to social psychologist Adams (1965), people expect their outcome of some exchange to be correlated[4] to inputs seen as relevant for that exchange, such as effort, skills and talent. Which inputs are considered relevant and how correlated individuals wish these inputs to be to the outcome should according to Adams depend on societal norms that individuals learn by socialisation. Dworkin (1981a,b), and later

* Corresponding author.
E-mail addresses: Ann-Sofie.Isaksson@economics.gu.se (A.-S. Isaksson), Annika.Lindskog@economics.gu.se (A. Lindskog).

[1] See for example Fong et al. (2005).

[2] See for example Burrows and Loomes (1994), Cappelen et al. (2007) and Clark (1998).

[3] See also the influential inequality aversion model of Fehr and Schmidt (1999), or fairness concepts stressing basic needs. See Konow (2003) for a good discussion of different fairness ideals.

[4] Interpreting Adam's equity theory in a strict sense, outcomes should even be *proportional* to inputs. For experimental evidence on this theme, see for example Van Dijk and Wilke (1994) or Clark (1998).

0167-2681/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2009.08.006

Roemer (2002), distinguish between inputs for which the individual could be considered directly responsible – 'responsible inputs', and those that are beyond the individual's control – 'arbitrary inputs', and argue that fair distributions should be based on responsible inputs only. If people in their fairness judgements actually distinguish between inputs in this fashion, then those who believe that income determinants to a greater degree are 'responsible' should consider the prevailing income distribution fairer and thus be less inclined to support redistribution, whereas those who to a larger extent view them as 'arbitrary' should see the existing income differences as more unfair and accordingly be more supportive of redistribution.[5]

With respect to empirical estimation of redistributive preferences, these arguments first of all motivate going beyond standard economic self-interest explanations. More specifically, they point to the importance of incorporating individual beliefs about the causes of income differences, and in particular beliefs on income determinants that could be seen as being under varying degrees of individual control. Second, they highlight the importance of studying preferences for redistribution in a country-comparative framework. Whether or not due to actual variation in what determines final incomes, beliefs about the causes of income differences are likely to vary across societies.[6] This should create corresponding differences in redistribution support. Similarly, judgements on the extent to which perceived income determinants are under individual control are likely to vary among individuals and communities. This variation too could be due to differences in norms as well as in actual circumstances. Regardless of which, the implication is that the relationship between beliefs about the causes of income differences and redistributive preferences is likely to vary with context, and not the least across countries, thus highlighting the importance of allowing for cross-country parameter heterogeneity.[7]

Against this background, this paper seeks to explain within- and between-country variation in redistributive preferences in terms of both self-interest concerns and an input-based concept of fairness captured by beliefs about the causes of income differences, allowing the effect of beliefs to differ among countries. More specifically, we will address the following two hypotheses:

Hypothesis 1. Both economic self-interest and an input-based fairness concept, where individuals judge the fairness of income determinants according to their perceived degree of 'responsibility', matter for redistributive preferences.

Hypothesis 2. Differences in beliefs about income determinants and differences in the effects of these beliefs both contribute to explain the cross-country variation in preferences for redistribution.

Several papers demonstrate reasons why a person's redistributive preferences do not necessarily correspond to his or her current pecuniary interest. Perceived prospects of future upward mobility and risks of future downward mobility may imply that a poor person sees redistribution as against her interest and that a rich person sees it as favouring her interests (Buchanan and Tullock, 1962; Piketty, 1995; Bénabou and Ok, 2001; Ravallion and Lokshin, 2000). Views on the incentive costs of redistribution are also likely to influence preferences for redistribution (Piketty, 1995). Moreover, Corneo and Grüner (2000) show, theoretically and empirically, that social competition and status concerns can have important influences on preferences for redistribution, and make the middle class align with the higher class to limit redistribution.

Explicitly relating beliefs about the causes of income differences to redistributive preferences is, however, a relatively new approach in the economics literature. Out of the few previous investigations, our study mostly resembles that of Fong (2001), who to our knowledge is the only one to explicitly distinguish between responsible and arbitrary inputs.[8] She examines a US sample and finds beliefs about causes of income differences to be important (and working in the expected directions) for explaining redistributive preferences. A few other studies also lend support to the importance of an input-based concept of fairness for redistributive preferences. Alesina and La Ferrara (2005) and Piketty (1995), for example, both confirm that in the US, those who believe that society offers equal opportunities to people who put in effort are more averse to redistribution.

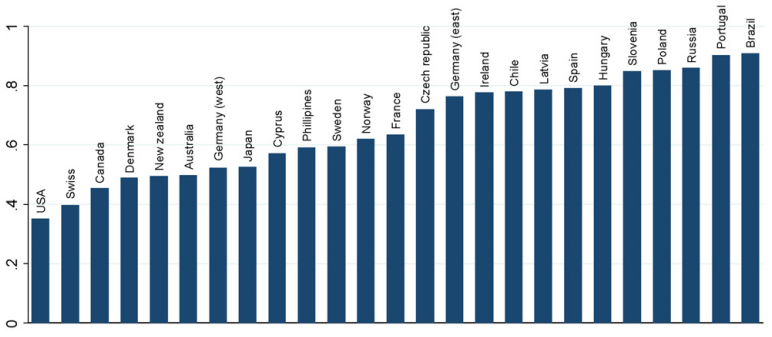
However, if there is country variation in beliefs about what causes income differences and in judgements of what income determinants could be considered under individual control one cannot necessarily expect these results to hold outside of the United States. In fact, related research efforts call attention to the need for cross-country-comparative work in the area. Based on a comparison of former East and West Germany showing that former East Germans are more in favour of redistribution than West Germans, even when controlling for their lower incomes, Alesina and Fuchs-Schündeln (2007) argue that individuals' preferences concerning government welfare provision are shaped by the economic regime in which they live. Similarly, Alesina et al. (2001) dedicate an extensive article to the issue of why the US does not have the same type of welfare state as Europe, and their evaluation does not speak to the advantage of basing conclusions of general attitudes towards redistribution on US evidence only.

[5] Cappelen and Tungodden (2006) add some nuance to this general claim, showing that if there are negative correlations between different non-responsibility (what we refer to as arbitrary) factors, one cannot expect a monotonic relationship between the responsibility assigned to people and the ideal level of redistribution. However, the general formulation put forward here should still hold.

[6] One reason for this variation could be differences in the redistributive policies pursued by the countries in question, in turn giving country variation in perceived and/or actual income earning possibilities (see for example the article on multiple welfare states equilibriums by Alesina and Angeletos, 2005). Whether country differences in beliefs about the causes of income differences are due to actual variation in what determines final incomes is an interesting question, but will not be addressed in this paper. Focus lies on the effect of these beliefs on redistribution support, rather than on their formation.

[7] The relationship between beliefs about the causes of income differences and redistributive preferences could vary between countries for several reasons, something which we get back to in Section 3.2.2.

[8] She refers to them as exogenous and discretionary factors.



[1] *'It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes'*

Fig. 1. Share of respondents agreeing or strongly agreeing with the redistributive statement[1].

In spite of these concerns, the country-comparative material relating redistributive preferences to beliefs about the causes of income differences is meagre. To our knowledge, the only serious study in the area based on a more than one-country sample is the paper by Corneo and Grüner (2002), which looks at 12 countries. Their main focus is the effect of social rivalry and status concerns on preferences for redistribution.[9] However, they also consider, and find a significant effect of, beliefs about the importance of hard work for determining income, and that people in former socialist countries are more supportive of redistribution. However, they do not, as is done in this paper, include variables capturing beliefs on income determinants that could be seen as being under a varying degree of individual control, nor is their approach country comparative in the sense that it allows for cross-country parameter heterogeneity.

This paper thus contributes to the literature by explicitly relating redistributive preferences to beliefs about income determinants under a varying degree of individual responsibility, and by doing so in a country-comparative framework seeking to explain both within-country and between-country variations.

2. Empirical framework

To investigate how preferences for redistribution vary within and between countries we use the ISSP Social Inequality III survey data set from 1999/2000 for 24 countries; Australia, Brazil, Canada, Chile, Cyprus, Czech Republic, Denmark, France, Germany, Hungary, Ireland, Japan, Latvia, New Zealand, Norway, Philippines, Poland, Portugal, Russia, Slovenia, Spain, Sweden, Switzerland, and USA.[10] Since preferences for redistribution have been shown to vary between former East and West Germany (Alesina and Fuchs-Schündeln, 2007), Germany has been divided into its former East and West German regions, giving us an effective sample of 25 countries. In most countries we have an estimation sample of 600 to 1000 observations. The smallest samples are those of eastern and western Germany, with 309 and 511 observations. The largest samples are the French and Brazilian ones, with 1396 and 1327 observations. In the total sample we have 20,250 respondents.

Our dependent variable is the response to the statement, '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*', ranging from 1 for strongly disagree to 5 for strongly agree. In using this variable as our dependent, we have to make the assumption that the responses to the statement actually reflect the degree of redistribution that the respondents want, meaning that people who are more supportive of the statement also desire more redistribution. The fact that responses to this statement are highly correlated with responses to a question about the desired progressiveness or regressiveness of the tax system makes us more confident with regard to this assumption.[11] As can be seen in Fig. 1, which gives the share of respondents agreeing or strongly agreeing with the redistributive statement in the country sub-samples, there is substantial country variation in support for redistribution. The share of respondents supporting the redistributive statement ranges from 34 percent in the US to around 91 percent in Brazil.

[9] Their key finding is that a person is more likely to favour redistribution if people that are either somewhat richer or somewhat poorer than them have a higher job status in relation to their own.

[10] Austria, Bulgaria, Great Britain, Israel, Netherlands, and Northern Ireland are excluded since key variables are missing.

[11] The reason why we do not use the tax question as our dependent variable is the much smaller variation over the five response categories for this question. Extremely few want high income earners to pay a smaller or much smaller share in taxes than low income earners, and these alternatives constitute two of the five response categories.

Turning to our explanatory variables, these could be divided into three major categories: pecuniary self-interest variables, indicators on beliefs about the causes of income differences included to capture the potential influence of input-based fairness concerns, and control variables. With regard to the former, an individual should according to economic thinking want the level of redistribution that maximises the utility derived from his/her current and expected future income. With redistribution going from the 'rich' to the 'poor', support for redistribution should thus be decreasing in both current and expected future relative income. Moreover, it is possible to view redistribution as insurance against income risk. A more risk-averse person should then prefer more redistribution and vice versa, and similarly someone with a high perceived income risk should prefer more redistribution and vice versa. Due to data limitations, however, expected future income, risk-aversion and perceived income risk are omitted; leaving us with relative income[12] to capture self-interest.

Other socio-demographic variables, such as class affiliation and higher education, could also be seen as considered to capturing self-interest, but might just as well capture differences associated with fairness concerns. Just as a more homogenous group is likely to be more equal in terms of omitted self-interest variables (such as expected future income), it seems reasonable that they also have more similar beliefs about how much an omitted 'input' does and should contribute to income. This ambiguity makes it more suitable to view the included socio-demographic indicators as controls for omitted variables rather than as factors in themselves capturing the influence of either fairness or self-interest concerns. The socio-demographic variables included on top of relative income are level of education, father's education, self-reported class belonging, sex and age.

Furthermore, we include a dummy controlling for potential concerns over incentives effects of redistribution (indicating whether the respondent agrees with the statement, '*Large differences in income are necessary for [country's] prosperity*'.[13] The pooled sample estimations also include country dummies to capture unexplained country differences in redistribution support.

To evaluate the potential influence of an input-based fairness concept where individuals judge the fairness of income determinants according to their perceived degree of 'responsibility', we need to include variables capturing beliefs about the importance of income determinants that are arguably under a varying degree of individual control.[14] As noted, views on the degree to which an input could be seen as 'responsible' are likely to differ among individuals. Some inputs, however, are easier to classify than others. Effort, for example, is often put forward as being largely under individual control, whereas factors associated with birth conditions, such as family background, could hardly be seen as something controllable by the individual. Inputs such as intelligence, skills or talents seem to be more controversial. We include three variables to capture beliefs about the importance of certain factors for determining income differences in society: one looks at beliefs about the importance of effort (arguably a responsible factor), another has to do with the importance of family background (arguably an arbitrary factor outside of individual control), and the third captures the perceived importance of intelligence and skills.[15] How to categorise the latter in terms of 'responsibility' is less clear-cut,[16] why the impact of this belief variable on redistributive preferences should be equally ambiguous and thereby occupy a middle position between the effects of the other two belief variables. For variable definitions see Table A1.

Since our dependent variable is discrete and inherently ordered, we use ordered probit for estimation according to the benchmark setup given in Eq. (1):

$$PR_{ic} = \alpha_c \tilde{y}_{ic} + \boldsymbol{\beta}_c' \mathbf{b}_{ic} + \boldsymbol{\delta}_c' \mathbf{x}_{ic} + \varepsilon_{ic} \tag{1}$$

PR_{ic} gives the unobserved redistributive preference of individual i in country c, $\tilde{y}_{ic}$ captures individual relative income, $\mathbf{b}_{ic}$ is the vector of belief PR_{ic} variables, $\mathbf{x}_{ic}$ is the vector of control variables, and ε_{ic} is a standard normally distributed error term. The probability that individual i in country c chooses response alternative k is the probability that the value of the unobserved support for redistribution falls between the cut-points μ_{k-1} and μ_k. Assuming normally distributed error terms

[12] Household income per adult equivalent divided by the country sample average. Note that the difference between relative income and absolute income is only relevant in pooled sample estimations including all four countries.

[13] It would be possible to follow Corneo and Grüner (2002) and control for the effect of status concerns on preferences for redistribution. Following their approach would, however, involve dropping observations from the richest and poorest income group, decreasing the representativeness of our sample and the variation in a key variable. Since status concerns is not our focus we choose not to do this.

[14] Some authors make a clear distinction between arbitrary and responsible inputs (see for example Cappelen and Tungodden, 2006, who refer to a strict 'responsibility cut'). We believe that speaking in terms of different *degrees* of responsibility over inputs, where completely arbitrary and entirely responsible are the two extremes, better reflect popular opinions in this context.

[15] The belief variables are based on questions asking how important the factor is 'for getting ahead', or on agreement with a statement saying that the factor is 'rewarded' in society (see Table A1). Although these formulations could be interpreted in non-monetary terms, we still believe that the answers constitute good approximations of beliefs about factors underlying *monetary* success. Hence we speak of these variables as concerning beliefs about the causes of income/income differences.

[16] Adding to this ambiguity is the dubious nature of the variable formulation. The statement captures both intelligence *and* skills, and many might argue that these two characteristics vary in terms of the extent to which they are acquired through life and thereby in the degree to which they are under individual control.

with mean zero and variance 1, and denoting the normal cumulative distribution function Φ, these are:

$$\begin{aligned} Pr(y_{ic}=1) &= \Phi(\mu_1 - \alpha_c\tilde{y}_{ic} - \boldsymbol{\beta}_c'\mathbf{b}_{ic} - \boldsymbol{\delta}_c'\mathbf{x}_{ic}),\\ Pr(y_{ic}=k) &= \Phi(\mu_k - \alpha_c\tilde{y}_{ic} - \boldsymbol{\beta}_c'\mathbf{b}_{ic} - \boldsymbol{\delta}_c'\mathbf{x}_{ic}) - \Phi(\mu_{k-1} - \alpha_c\tilde{y}_{ic} - \boldsymbol{\beta}_c'\mathbf{b}_{ic} - \boldsymbol{\delta}_c'\mathbf{x}_{ic}),\\ Pr(y_{ic}=5) &= 1 - \Phi(\mu_4 - \alpha_c\tilde{y}_{ic} - \boldsymbol{\beta}_c'\mathbf{b}_{ic} - \boldsymbol{\delta}_c'\mathbf{x}_{ic}), \quad k=2,3,4. \end{aligned} \tag{2}$$

Regression coefficients and cut-points are estimated by the maximum likelihood estimator. When the coefficient is positive, a positive change in the independent variable decreases the probability of the lowest ranked outcome and increases the probability of the highest ranked outcome, but does not reveal the direction of change in probabilities of intermediate outcomes. To be able to say something about the direction of change for intermediate outcomes, as well as of magnitudes of changes, we present the effects of given discrete changes in the independent variables on the probabilities of observing the different outcomes on our dependent variable. For a dummy variable D this is simply calculated as $\Pr(y_{ic}=k)$ evaluated at $D=1$ minus $\Pr(y_{ic}=k)$ evaluated at $D=0$, keeping the remaining variables at their means. Analogously, for the effect of a given change in a continuous variable X, $\Pr(y_{ic}=k)$ is evaluated at two specified values of X.

3. Results

In this section we evaluate our two hypotheses empirically. We start by examining our first hypothesis, considering the extent to which economic self-interest considerations and input-based fairness concerns can help explain redistributive preferences. Then we turn to our second hypothesis, suggesting that both differences in beliefs about income determinants and differences in the effects of these beliefs contribute to explain the cross-country variation in redistributive preferences.

3.1. Explaining preferences for redistribution

Our first hypothesis can be evaluated by considering the results of the benchmark estimation given in Eq. (1), estimated separately for each country as well as for the full sample with country dummies. The first part of this hypothesis, stipulating that self-interest considerations should matter for redistributive preferences, implies that a higher relative income should give a lower support for redistribution, so that $\alpha_c < 0$. The analysis of the second part of the hypothesis, arguing that the effect of beliefs about the causes of income differences differs with the respective inputs' degree of responsibility, rests on accepting the suggested classification of effort as the most 'responsible' input out of the three considered, family background as the least responsible, and intelligence/skills as a less clear-cut one located somewhere between the other two. Then with regard to believing that the concerned inputs are important for determining income, we should have $\beta_c^{effort} < \beta_c^{skills} < \beta_c^{family}$, $\beta_c^{effort} < 0$ and $\beta_c^{family} > 0$.

Table 1 presents the marginal effects from the pooled sample ordered probit estimation of Eq. (1). Figs. 2–5 summarise the results of estimation of Eq. (1) for our 25 country sub-samples, focusing on the effects of movements in our key variables – the belief and relative income indicators.

Let us start by briefly commenting on the pooled sample effects of our control variables (Table 1). As noted in Section 2, omitted belief and self-interest variables makes the parameters of the socio-demographic controls somewhat difficult to interpret; do they reflect differences in norms and beliefs among different groups in society, or do they capture self-interest considerations? At any rate, a number of interesting patterns stand out. Respondents with higher education, respondents whose fathers have higher education, and respondents claiming to belong to the upper class all tend to be less supportive of redistribution (the reverse is true for those who claim to belong to the working class). This could reflect higher expected future relative incomes given current relative income for well-educated people with steeper age-earnings profiles, or that privileged classes have better professional connections and thus face smaller income risks, but could also depend on differences in norms between social groups. Similarly, the fact that women are more likely to support the redistributive statement could reflect a higher perceived income risk among women, a greater degree of risk-aversion or alternatively that women hold different norms regarding what is fair. Moreover, there is a positive age effect on support for the redistributive statement, perhaps reflecting a change over time in popular sentiments towards redistribution. In what follows these socio-demographic variables will be treated as controls for omitted self-interest and belief indicators. The estimation also includes 24 country dummies, where USA, the country with the least support for redistribution, is the reference category. The country effects, all statistically significant at the 1 percent level, are not presented in Table 1, but to get a feeling for their size see Fig. 9, Specification 2. In short, the largest country effects are found in the former socialist countries (in line with the findings of Corneo and Grüner, 2002), in four countries with a recent history of right-wing authoritarian regimes – Brazil, Chile, Portugal, and Spain – and in France.[17] Finally, and as expected, people who claim inequality is needed for prosperity – a variable included to control for concerns about possible incentive effects of redistribution – are less supportive of redistribution.

Turning to our self-interest variable, the results of the pooled sample estimation indicate that, as expected, people with a higher relative income tend to be less supportive of redistribution. However, the effect is quite small. Conditional on our

[17] Log-likelihood ratio tests show that the class variables, as well as the country dummies, are jointly important (the test statistics are 274.44 and 2533.11, giving p-values at 0.000).

Table 1
Pooled sample marginal effects[a] on probability of agreeing with the redistributive statement.[b].

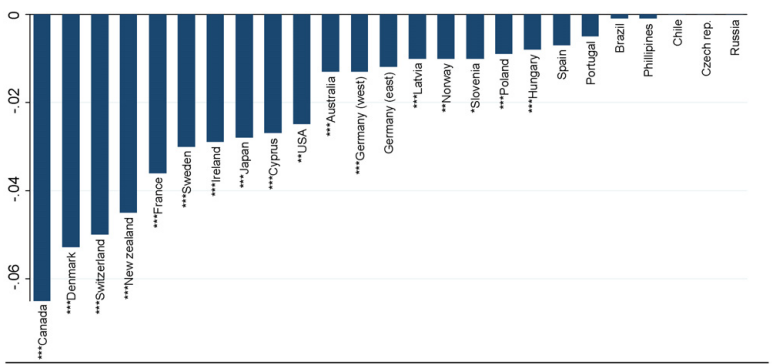
	Marginal effect on choosing response category				
	Strongly disagree	Disagree	Neither	Agree	Strongly agree
Believe wealthy family important	−0.018***	−0.035***	−0.029***	−0.003***	0.086***
	0.001	0.003	0.003	0.001	0.007
No strong family belief	−0.002	−0.004	−0.003	0.000	0.009
	0.001	0.003	0.002	0.000	0.006
Believe effort rewarded	0.013***	0.024***	0.019***	−0.002***	−0.054***
	0.002	0.004	0.003	0.001	0.008
No strong effort belief	0.011***	0.019***	0.015***	−0.002***	−0.043***
	0.002	0.004	0.003	0.001	0.008
Believe intell./skills rewarded	0.006***	0.011***	0.008***	0.000	−0.024***
	0.002	0.003	0.003	0.000	0.008
No strong intell./skill belief	0.005**	0.009**	0.007**	−0.001	−0.020**
	0.002	0.004	0.003	0.000	0.008
Relative income	0.002***	0.004***	0.003***	0.000	−0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Age	−0.001***	−0.002***	−0.002***	0.000	0.006***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
Female	−0.012***	−0.022***	−0.017***	0.000	0.051***
	(0.001)	(0.002)	(0.002)	(0.000)	(0.005)
Higher education	0.012***	0.022***	0.017***	−0.002***	−0.048***
	(0.002)	(0.003)	(0.002)	(0.001)	(0.007)
Father has higher education	0.010***	0.018***	0.014***	−0.002***	−0.041***
	(0.002)	(0.003)	(0.002)	(0.001)	(0.007)
Upper class	0.027***	0.043***	0.031***	−0.011***	−0.090***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.008)
Working class	−0.016***	−0.030***	−0.025***	−0.002**	0.074***
	(0.001)	(0.003)	(0.002)	(0.001)	(0.007)
Inequality necessary for prosperity	0.015***	0.026***	0.020***	−0.003***	−0.058***
	(0.002)	(0.003)	(0.002)	(0.001)	(0.006)
The estimation also includes 24 country dummies.					
Observations	20250				

[a] Dummy variable effects (all except for relative income and age) are for a discrete 0-1 change (for reference categories, see Table A1). The relative income effect is for a one median absolute deviation increase around the median, and the age effect is for an increase from 40 to 50 (approximately equivalent to a one standard deviation change around the mean).

[b] '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*'.

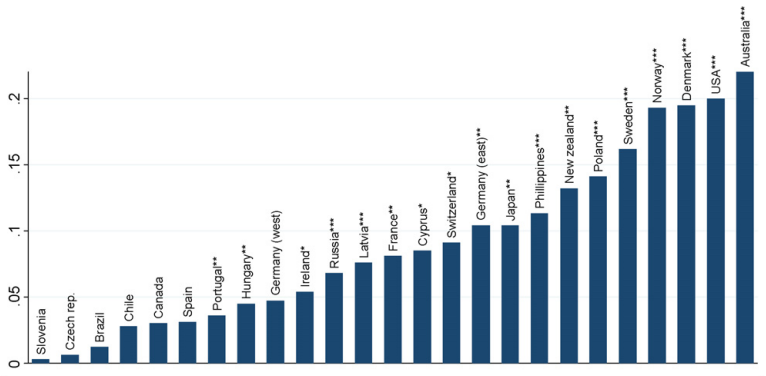
** Significance at the 5% level.

*** Significance at the 1% level.



Based on ordered probit estimation of equation 1 for the country sub-samples. Presents the effects of a one median absolute deviation increase around the median in relative income on the probability of agreeing or strongly agreeing with the statement, '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*'.

Fig. 2. Country variation in the effects of a higher relative income on support for redistribution.

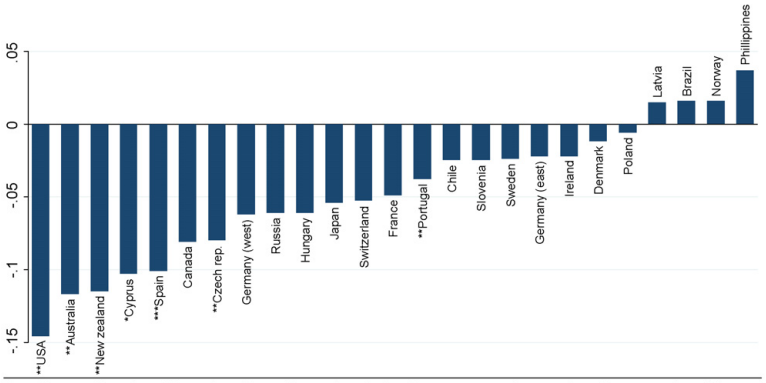


Based on ordered probit estimation of equation 1 for the country sub-samples. Presents the effects of believing it to be 'essential' or 'very important', rather than 'not very important' or 'not important at all', to come from a wealthy family to get ahead on the probability of agreeing or strongly agreeing with statement, '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*'.

Fig. 3. Country variation in the effects of holding the belief 'coming from a wealthy family is important to get ahead' on support for redistribution.

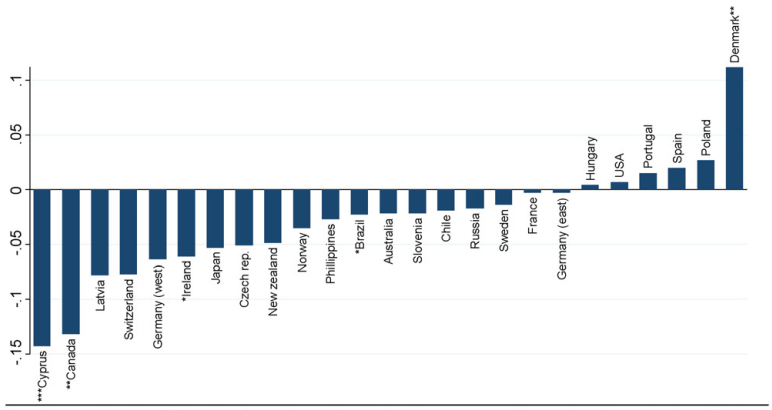
belief- and socio-demographic variables, an increase in relative income from one-half median absolute deviation below the median to one-half median absolute deviation above the median involves a one percentage point reduction in the probability of supporting the redistributive statement. Considering the relative income effects in the individual country sub-samples (Fig. 2), in the great majority of countries a higher relative income is associated with a statistically significant smaller probability to support redistribution. The associations are far from homogenous, however. In Canada, the concerned relative income change involves a 7 percentage point smaller probability to support redistribution. In Hungary the equivalent reduction is around 1 percentage point, i.e. in line with our pooled sample estimate. In 8 countries the relative income effect is not statistically different from zero.

Turning to the effects of holding certain beliefs about what causes income differences, the pooled sample estimates (Table 1) indicate that the belief effects follow the hypothesised pattern. Believing that coming from a wealthy family is



Based on ordered probit estimation of equation 1 for the country sub-samples. Presents the effects of agreeing or strongly agreeing rather than disagreeing or strongly disagreeing to the statement '*In [country] people get rewarded for their effort*' on the probability of agreeing or strongly agreeing with statement, '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*'.

Fig. 4. Country variation in the effects of believing 'effort is rewarded' on support for redistribution.



Based on ordered probit estimation of equation 1 for the country sub-samples. Presents the effects of agreeing or strongly agreeing rather than disagreeing or strongly disagreeing to the statement *'In [country] people get rewarded for their intelligence and skills'* on the probability of agreeing or strongly agreeing with statement, *'It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes'.*

Fig. 5. Country variation in the effects of believing 'intelligence/skills is rewarded' on support for redistribution.

important to get ahead (the reference categories capture believing that the respective inputs are not important/rewarded, see Table A1) is, as anticipated, associated with stronger support for redistribution. In the pooled sample, it involves an 8 percentage point increase in the probability of supporting the redistributive statement, to be compared with the 1 percentage point reduction in the same probability associated with the relative income change. Also in line with our hypothesis, believing that effort is rewarded comes with a 6 percentage point lower probability to support the redistributive statement. Similarly, believing that intelligence/skills are rewarded involves a 2 percentage point lower probability to support the statement. Hence, the pooled sample belief effects vary according to the pattern ($\beta_c^{effort} < \beta_c^{skills} < \beta_c^{family}$), $\beta_c^{effort} < 0$ and $\beta_c^{family} > 0$ suggested by the respective inputs' degree of responsibility. However, looking at the individual country sub-samples reveals considerable heterogeneity.

In the majority of our country sub-samples believing that coming from a wealthy family is important to get ahead is associated with a statistically significant higher probability to support redistribution (Fig. 3). The size of the effect varies across countries, however; holding the belief involves a 4 percentage point higher probability to support redistribution in Portugal and a 22 percentage point higher probability to do so in Australia.

The effects of believing effort is rewarded (Fig. 4) are somewhat less clear-cut, but all statistically significant effects (seven countries) are negative – the decrease ranging from 4 percentage points in Portugal to 15 in the US.

As hypothesised, the effects of beliefs on rewards to intelligence and skills are most ambiguous. As seen in Fig. 5 believing that intelligence/skills is rewarded involves a statistically significant (2–14 percentage point) lower probability to support redistribution in Cyprus, Canada, Ireland and Brazil, and an 11 percentage point *higher* probability to do so in Denmark.

To formally test the joint importance of the belief variables, we perform log-likelihood ratio tests where the unrestricted model includes them and the restricted model does not (see Table A2, Panel 1). The null-hypothesis, that excluding the belief variables does not affect the explanatory power of the model, can be firmly rejected in the absolute majority of country sub-samples. Only in three countries (Chile, Slovenia and Switzerland) can we not reject the null.

To test not only if the belief variables matter, but if they do so in line with the pattern expected from the respective inputs degree of 'responsibility', we perform a number of one-sided tests to evaluate if the parameters follow the hypothesised pattern $\beta_c^{effort} < \beta_c^{skills} < \beta_c^{family}$, $\beta_c^{effort} < 0$ and $\beta_c^{family} > 0$ (see Table A3). The pooled sample tests confirm the hypothesised pattern. In the individual countries the picture is somewhat more mixed. Our hypothesis that the effect of believing family background to be rewarded has a positive effect on support for redistribution, and that this is larger than the effects of effort and intelligence/skills, is supported in the absolute majority of the country sub-samples. Our hypothesis that believing effort to be rewarded has a negative effect on redistribution support is confirmed in 12 of our country samples. For both family and effort, where we cannot reject the null we cannot reject the alternative hypothesis either. We find least support for the hypothesis that the intelligence/skills effect is larger than the effort effect, which is confirmed only in four

countries. In two countries the test actually indicates the reverse, that the effort effect is larger than the intelligence/skills effect.[18]

3.1.1. Omitted variables

When interpreting the results one has to consider the potential influence of omitted self-interest and belief variables on our key parameters. Variables that appear important in this context include the self-interest indicators expected future relative income, risk-aversion and perceived income risks, and variables capturing beliefs regarding the importance of a wide range of inputs which could affect income, for example luck, ethnicity and gender. Since self-interest indicators and beliefs about the causes of income are likely to vary among social groups, the included socio-demographic variables should pick up much of this variation, thus helping to alleviate the problem. Nevertheless, the issue deserves some attention.

First, the relative income estimate may be biased by omitted self-interest variables. For example, expected future income should be positively correlated with current relative income. If we assume that support for redistribution depends on some weighted average of current and expected future income, then the estimated relative income coefficient will be larger than its true effect as it also captures some of the effects from expected future income.

Another potential concern is if omitted belief variables are correlated with relative income, which, if we are interested in isolating the effect of relative income that is due to direct self-interest concerns, could bias the estimated relative income effect. Similarly, omitted self-interest variables could bias the estimated effects of our belief variables.[19] Again, if omitted beliefs and self-interest variables vary across social groups the socio-demographic variables should pick up much of this unobserved variation. For what remains, we naturally cannot investigate the covariations between relative income and omitted belief variables, or between omitted self-interest variables and our belief variables. What we can do is to evaluate the correlation pattern between relative income and our included belief variables, hoping that the latter reveals something about the former; if there is little correlation between included belief variables and relative income it seems less likely that we have a problem of correlation between omitted belief variables and relative income, or between omitted self-interest variables and our included belief variables. As it turns out, there is very low correlation between our relative income indicator and our belief variables. In the pooled sample, correlation coefficients (in absolute terms) range between 0.001 and 0.047. Moreover, testing for multicollinearity of regressors using variance inflation factors and the condition index shows that neither in the pooled sample nor in the country sub-samples do we have a problem of multicollinearity (the variance inflation factors are in the order of 1–2 and the condition numbers range from 2 to 5).

Still, to get a picture of whether relative income affects the beliefs regarding income determinants we run ordered probit regressions with the belief indicators as dependent variables and with relative income and the socio-demographic controls as independent variables (see Tables A4–A6), for the pooled sample as well as the individual country sub-samples. In the pooled sample there is actually a relative income effect on our family and effort variables. However, the effect is very small; a relative income increase of one median absolute deviation around the median involves a 0.3 percentage point smaller probability to believe that coming from a wealthy family is very important or essential to get ahead, and a 0.2 percentage point smaller probability to agree or strongly agree that effort is rewarded. Moreover, in the absolute majority of country sub-samples the effect of relative income is far from statistically significant. For none of the belief variables more than 4 out of 25 countries have significant relative income effects, and for the ones that do, the effects are again small.[20] If we exclude the countries (Brazil, France, Hungary, Latvia and Russia) where a statistically significant relative income effect is found for two belief variables and run a restricted pooled sample estimation, the relative income effect is no longer there, suggesting that the associations identified in the original pooled sample estimation were driven by a small number of countries differing from the overall pattern rather than by increased precision following from more observations. Hence, with the exception of a few countries, we identify no effect of relative income on our belief variables. If the same goes for the omitted belief variables, then their influence should not be a major problem.

An alternative approach could be to argue that the stability of the relative income effect to the inclusion of the belief variables might indicate whether omitted belief variables constitute a problem. Estimating our benchmark model (Eq. (1)) with and without the belief variables[21] it turns out that in the pooled sample as well as in all the country sub-samples the relative income parameter is very stable.

Summing up, problems of omitted variables make it difficult to pin down the exact magnitudes of the effects found. We can nevertheless conclude that on the whole, our relative income indicator appear to capture self-interest considerations and our belief variables fairness concerns. Our estimations suggest that both relative income and beliefs about the causes of income differences are important to explain redistributive preferences, and that they do so according to the pattern suggested in Hypothesis 1. We can, at this stage, also note that there is substantial country variation in redistributive preferences, as well as in the effects of our main explanatory variables on these. In the next section we investigate this variation further.

[18] The tests of the alternative hypotheses are not presented, but are available from the authors.

[19] Of course, omitted beliefs could also bias the estimated effects of the included beliefs. We see this as less of a problem, since then we can assign the effects of belief variables to fairness considerations rather than to self-interest concerns.

[20] A relative income increase of one median absolute deviation around the median in most cases involves a smaller than one (and never more than 1.7) percentage point change in the probability to support the concerned statements.

[21] The results are available from the authors.

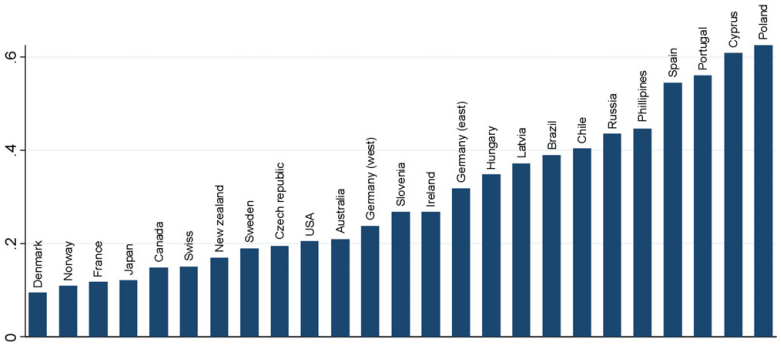


Fig. 6. Share of respondents believing that it is essential or very important to be from a wealthy family to get ahead.

3.2. Explaining country variation in redistributive preferences

Our second hypothesis stipulates that differences in beliefs about the causes of income differences, as well as differences in the *effects* of these beliefs, contribute to explain the cross-country variation in redistributive preferences. We will evaluate this hypothesis in three steps. First, we consider whether beliefs about the causes of income differences differ across countries in a direction consistent with the country variation in redistributive support. Second, we examine whether there is cross-country heterogeneity in the effects of holding certain beliefs regarding what causes income differences on redistributive preferences. Finally, we bring the picture together by addressing the extent to which the discussed differences in beliefs and impacts of these beliefs can explain the observed country variation in redistributive preferences.

3.2.1. Country differences in beliefs about income determinants

Let us start by considering the country variation in beliefs about what causes income differences. Figs. 6–8 give the country shares of respondents who believe that coming from a wealthy family is important to get ahead, or agree to that effort and intelligence/skills are rewarded in the country. As expected, there is substantial country variation in beliefs about income determinants. The share of respondents who believe that it is important to be from a wealthy family to get ahead ranges from 8 percent in Denmark to 62 percent in Poland. For the beliefs about whether effort and intelligence/skills are rewarded, the lowest shares of respondents believing so are found in Russia, where 8 and 10 percent agree with the respective statements, and the highest in the US, where the equivalent shares are 67 and 75 percent.

The above shares give an overview of the country variation in beliefs about the causes of income differences, but do not inform us about the full variation in the belief distributions. To formally test whether the distributions of beliefs differ

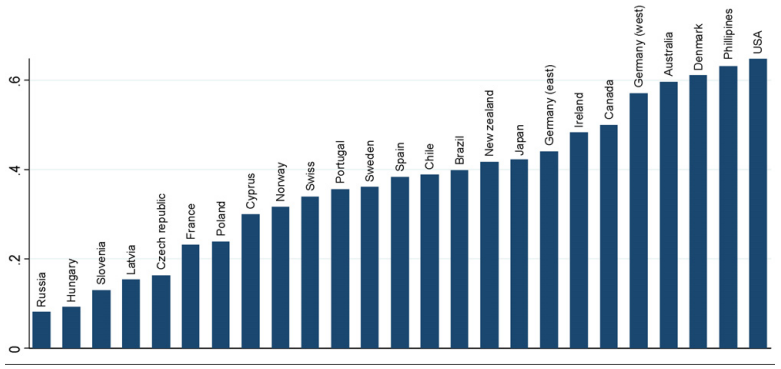


Fig. 7. Share of respondents agreeing or strongly agreeing to the statement 'In [country] people get rewarded for their effort'.

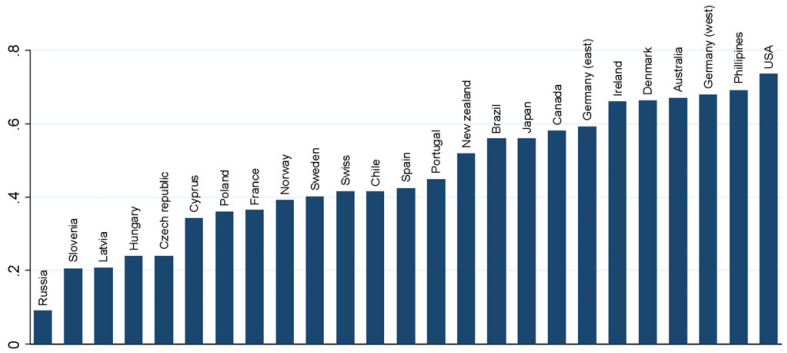


Fig. 8. Share of respondents agreeing or strongly agreeing to the statement 'In [country] people get rewarded for intelligence and skills'.

between countries we perform two-sample Kolmogorov–Smirnov tests of equal cumulative distribution functions.[22] We compare each country with the remaining countries for the three belief variables, resulting in 975 tests. The null-hypothesis of equal cumulative distribution functions was rejected at the five percent level of significance in 938 tests. To draw any conclusions about similarity of distributions in the special cases where we cannot reject the null of equal distributions, a test with an exact *p*-value would be necessary. Overall it seems fair to say that beliefs about the causes of income differences vary across countries.

With effort being classified as the most and family background as the least responsible input, one would predict that the countries that to a greater extent believe that effort is rewarded in society and that family background is not very important for getting ahead will also be the ones least supportive of redistribution (and vice versa). Inspection of Fig. 1 and Figs. 6–8 lends some support to this hypothesis. USA, the country where believing that effort is rewarded is most common (Fig. 7), is also the country with the least support for redistribution (Fig. 1). Other countries where the belief that effort is rewarded is common are the Philippines, Denmark, Australia, Germany (west) and Canada, all of which have comparatively low support for redistribution. Respondents in the former socialist countries in Central and Eastern Europe are generally very sceptical of the claim that effort is rewarded, while at the same time showing strong support for redistribution. A similar pattern can be observed for beliefs concerning the rewards to intelligence and skills (Fig. 8).

Believing that it is important to be from a wealthy family to get ahead (Fig. 6) is most common in Poland, Cyprus, Portugal and Spain, and at least Poland, Portugal and Spain display comparatively strong support for redistribution. Correspondingly, in countries where respondents do not believe that family background is very important for getting ahead – Denmark, Norway, France, Japan and Canada – the support for redistribution is comparatively low. At this stage it thus seems as though country differences in beliefs about income differences could have some relevance for explaining cross-country differences in redistribution support.

3.2.2. Country differences in the effects of beliefs about income determinants

Let us now turn to the second step, where we evaluate possible cross-country heterogeneity in the *effects* of the belief variables on redistributive preferences. Figs. 3–5, presenting the effects of belief variables on redistribution support in the respective countries, suggest such heterogeneity to be present. As noted, believing it to be important to be from a wealthy family to get ahead is for the great majority of countries associated with a higher probability to support redistribution. The largest effects of holding this belief are found in the US, Australia, Denmark and Norway, where it implies an approximate 20 percentage point increase in the probability of supporting the redistributive statement. However, in a few countries – Brazil, Canada, Chile, Czech Republic, Germany (west), Slovenia, and Spain – the effect is comparatively small and statistically insignificant. According to the reasoning in this paper, this fact could be interpreted as people from these countries assigning some degree of individual responsibility over family background. While it is difficult to argue that people can affect which family they are born into, the argument that someone who has succeeded in creating wealth should be able to pass this on to his/her children is quite common. The degree of responsibility assigned to an input may not necessarily depend only on perceived *individual* control over that input; conceivably it could also depend on perceived control within a larger entity, such as the family.[23] An alternative interpretation could be that in these countries people are more libertarian in the sense

[22] The Kolmogorov–Smirnov test is non-parametric and sensitive to differences in both the location of the distribution and the shape of the distribution. It is designed for testing the distribution of continuous variables, but has been demonstrated to be applicable to discrete random variables too, in which case it is conservative; i.e. for a given level of significance the null hypothesis of equal empirical distributions will be rejected less or as often as with the exact true test statistic (Conover, 1999; Goodman, 1954; Noether, 1963).

[23] See for example the theoretical model in Alesina and Angeletos (2005).

that they believe a person is entitled to the income he/she earns, irrespective of his/her degree of control over the inputs involved in earning that income.

In the US, believing effort to be rewarded implies an approximate 15 percentage point decrease in the probability of supporting the redistributive statement. In other 'Neo-European' countries – Australia, New Zealand, and Canada – the decrease is of around 10 percentage points. In the Southern European countries Cyprus, Spain, France, and Portugal, in the two former socialist countries Hungary and Czech Republic, and in Japan there are also statistically significant negative effort effects on the probability of agreement with the redistributive statement (ranging from 4 to 10 percentage points).[24] In other countries – the Scandinavian countries Denmark, Norway and Sweden, eastern and western Germany and Switzerland, the remaining four former socialist countries, and in Brazil, Chile, Ireland and the Philippines – however, the effects of believing that effort is rewarded are not statistically significant and do not stand out as large. This could be taken to indicate that in these countries effort is not to the same extent viewed as an input under individual control. Indeed, it is conceivable that depending on social background and other circumstances, individuals do not all have the same choice set regarding how much effort to exert. An alternative interpretation is that people in these countries are more concerned about equal outcomes, regardless of the degree of control they believe people have over important income determinants.

Believing intelligence and skills to be rewarded produces mixed results; in Denmark it implies an 11 percentage point *increase* in the probability of agreeing or strongly agreeing with the redistributive statement, whereas in Brazil, Canada, Cyprus, Latvia and the Czech Republic[25] it comes with a *decrease* in the same probability in the range of 2–14 percentage points. This could be taken to suggest country differences in the degree of responsibility assigned to this input, but could also indicate that the countries differ in the fairness ideals adhered to, with Danes being more concerned with equal outcomes and Brazilians, Canadians, Cypriots, Latvians and Czechs being more libertarian. In other countries the effect of believing intelligence and skills to be rewarded is not statistically different from zero.

We formally test whether the effects of belief variables differ across countries using a number of log-likelihood ratio tests (see Table A2, Panel 2). First, a restricted model in which country differences are only allowed to affect the intercept is firmly rejected in favour of a model that allows different slopes of the belief parameters, thus confirming the suspected presence of cross-country heterogeneity in the belief effects. Next, we test if there is parameter heterogeneity with respect to the beliefs regarding each input separately. For all inputs, the hypothesis of homogenous effects can be safely rejected.

3.2.3. Can the differences in beliefs and the differences in effects of these beliefs help explain cross-country variation in redistribution support?

Let us now turn to the last stage where we address to what extent the identified differences in (1) beliefs and (2) the impacts of these beliefs can explain the large country variation observed in redistributive preferences. Previous literature suggests that differences in people's beliefs are central in this respect (Alesina and Angeletos, 2005). To get an idea of the relative importance of differences in beliefs and differences in the *effects* of these beliefs for explaining cross-country variation in preferences for redistribution, we evaluate how the effect of belonging to a certain country changes as beliefs and beliefs-country interaction terms are added to the model. USA, the country hitherto most studied with regard to preferences for redistribution, and the country with the least support for redistribution controlling for other variables, is the reference country. To be more specific, we estimate the following three equations and focus on whether the parameters in $\boldsymbol{\varphi}$ approach zero as we allow for differences in beliefs (2) and differences in the effects of these beliefs (3).

$$1 \quad PR_{ic} = \boldsymbol{\phi}'\mathbf{country}_{ic} + \boldsymbol{\delta}_c'\mathbf{x}_{ic} + \varepsilon_{ic}$$

$$2 \quad PR_{ic} = \boldsymbol{\phi}'\mathbf{country}_{ic} + \boldsymbol{\beta}'\mathbf{b}_{ic} + \boldsymbol{\delta}_c'\mathbf{x}_{ic} + \varepsilon_{ic}$$

$$3 \quad PR_{ic} = \boldsymbol{\phi}'\mathbf{country}_{ic} + \boldsymbol{\beta}'\mathbf{b}_{ic} + \boldsymbol{\gamma}'\mathbf{b}_{ic}\mathbf{country}_{ic} + \boldsymbol{\delta}_c'x_{ic} + \varepsilon_{ic}.$$

Fig. 9 shows the resulting country effects.

Adding the belief variables to the model marginally reduces the effect of being of a certain nationality.[26] Furthermore, the country effect with beliefs (Specification 2) always falls within the 95 percent confidence interval around the country effect without beliefs (Specification 1).[27] Unexpectedly, differences in beliefs can thus, at most, explain a very small part of the cross-country variation in redistributive preferences.

Allowing for heterogeneity in the effects of beliefs reduces country effects quite substantially (the exceptions are eastern and western Germany where it increases the country effect marginally).[28] The effect of being Australian, Danish or Philippine (rather than American) almost disappears. The effect of being Canadian, Cypriot, Czech, Japanese, New Zealander, or Swedish

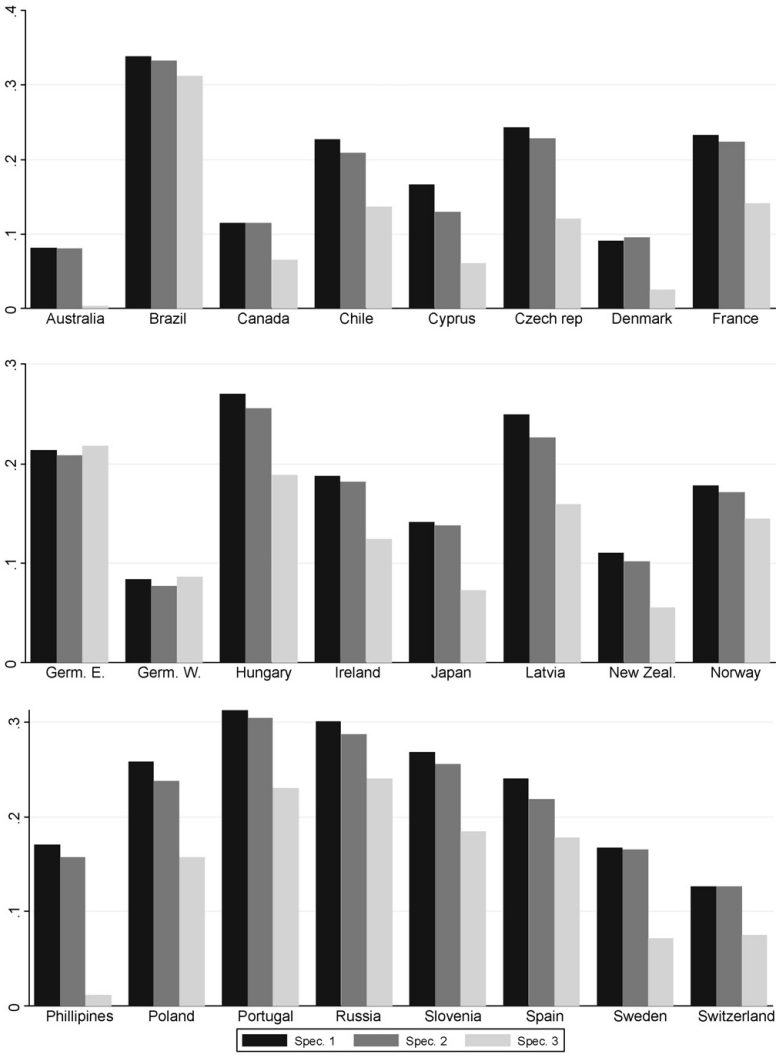
[24] For Canada, Cyprus, France, Hungary, and Japan a negative effect was confirmed using the one-sided test in Table A3, though the effect can only almost be rejected to equal zero at the 10 percent level of significance using a two sided test (Fig. 4).

[25] In the Czech sample it cannot be rejected at the 10 percent level that the effect differs from zero using a two-sided test, but using a one-sided test it can be rejected that the effect is equal to or larger than zero at the 5.5 level of significance.

[26] Comparing other countries with each other, the difference between their country effects (relative to the US) should not have been affected more than marginally either, since all changes in country effects are small and in the same direction.

[27] The estimated country effects and their 95% confidence intervals are available from the authors.

[28] Comparing other countries with each other, the difference between their country effects (relative to the US) increases in some cases, but the fact remains that country effects decrease vis-à-vis the extreme low redistribution support case USA.



Country dummy effects on the probability to agree or strongly agree with the redistributive statement, from ordered probit estimations with the following explanatory variables (see Table A1 for more detail):
Spec. 1: country dummies + other controls
Spec. 2: country dummies + other controls + belief variables
Spec. 3: country dummies + other controls + belief variables + belief*country

Fig. 9. Explaining country variation in redistributive preferences with differences in beliefs and differences in effects of beliefs.

is approximately halved. About one-third of the effect of being Chilean, French, Irish, Latvian, Polish, or Swiss disappears, and roughly one fourth of that of being Hungarian, Portuguese, or Slovenian.[29] Hence, as it seems, a relatively large part of many countries' stronger support for redistribution (compared to the US) could be explained by people in these countries assigning a lower degree of responsibility to inputs believed to be important for income determination. However, for some countries, such as Brazil and Russia, different effects of holding certain beliefs about income determinants on preferences for redistribution does little to explain their stronger support for redistribution, and the stronger support for redistribution in Germany than in the US becomes, if anything, even more puzzling, considering that Germans seemingly assign a higher degree of responsibility to inputs believed to determine income.

The conclusion we can draw from this is that while differences in beliefs on what causes income differences seem to be important for explaining within-country variation in redistributive preferences, they do little to explain between-country differences. Differences in the effects of holding certain beliefs, however, seem to be important for explaining between-country variation in redistributive preferences.

4. Conclusions

The objective of this study was to explain variation in redistributive preferences, within as well as between countries, in terms of self-interest concerns and an input-based concept of fairness. The latter was captured by the effect of beliefs about the causes of income differences. We included beliefs about income determinants arguably under varying degrees of individual control, stipulating that believing a 'responsible' factor to be important for determining income would imply less support for redistribution, whereas believing an input outside individual control to be an important income determinant should bring with it more support for redistribution. Importantly, we argued that these beliefs, and their effects, should vary with context. The country-comparative perspective was therefore central; we aimed to explain not only within-country but also between-country variation in redistributive preferences. Two hypotheses were formulated and tested using data for 25 countries.

Our first hypothesis suggested that both economic self-interest and an input-based fairness concept, where individuals judge the fairness of income determinants according to their perceived degree of 'responsibility', matter for redistributive preferences. This was supported by the data. In the pooled sample, relative income had a negative, although quite modest, impact on preferences for redistribution, and the effects of the variables capturing beliefs about the causes of income differences followed the hypothesised pattern $\beta_c^{effort} < \beta_c^{skills} < \beta_c^{family}$, $\beta_c^{effort} < 0$ and $\beta_c^{family} > 0$. As stipulated, believing effort (a 'responsible input') to be rewarded in society had a negative impact on support for redistribution, whereas believing that family background (an 'arbitrary input') is important for getting ahead was associated with stronger support for redistribution. Also, and as expected, the effect of believing that intelligence/skills (the input arguably most difficult to classify in terms of 'responsibility') are rewarded fell in-between those of believing family or effort to be important income determinants.

On the whole, the country sub-sample estimations supported our hypothesis, but revealed considerable heterogeneity in terms of the magnitude and statistical significance of effects. A higher relative income was generally associated with less support for redistribution, but the size of the effect varied. In the countries where we found statistically significant belief effects, the family effects were positive and the effort effects negative, albeit varying considerably in magnitude. Moreover, and as hypothesised, the family effect was larger than the effort and intelligence/skills effects. Comparing the magnitudes of the effort and intelligence/skills effects, however, the results were mixed – we could not establish that the effect of believing the responsible input effort to be important is smaller than that of believing the more ambiguous input intelligence/skills to be so.

Hence, whereas the aggregate pattern suggested that individuals base their preferences for redistribution on self-interest considerations as well as input-based fairness concerns, the country comparison revealed that this pattern is not necessarily universal.

Our second hypothesis put forward that differences in both beliefs about income determinants, and in the effects of these beliefs, should contribute to explain the cross-country variation in redistributive preferences. We could establish that there is considerable country variation in beliefs about income determinants, and that this variation often follows a pattern that would be expected judging from our input-based fairness concept and the observed country variation in redistribution support. Somewhat surprisingly, however, our analysis suggested that country differences in beliefs about income determinants, at best, could explain very little of the country variation in redistributive support.

Turning to the *effects* of the belief variables, our results showed that these too vary significantly across countries. Believing that coming from a wealthy family is important to get ahead in some countries had basically no effect while in others it involved an over 20 percentage point increase in the probability of supporting redistribution. The effect of believing effort to be an important income determinant varied from being statistically non-discernible from zero to decreasing the probability

[29] For eleven countries (Australia, Chile, Cyprus, Czech Republic, Denmark, France, Japan, Philippines, Poland, Portugal and Sweden) the country effect with beliefs parameter heterogeneity (Specification 3) falls outside of the 95 percent confidence interval of country effects with homogenous (Specification 2) or no (Specification 1) beliefs. For four other countries (Hungary, Latvia, Russia, Slovenia) it falls outside of the confidence interval of the country effect with no beliefs and at the limit of the confidence interval of that with homogenous belief parameters, and for nine countries (Brazil, Canada, Germany (east), Germany (west), Ireland, New Zealand, Norway, Spain, Switzerland) the country effect falls within the confidence intervals of the prior country effects.

of supporting redistribution with around 15 percentage points. With respect to believing intelligence/skills to be rewarded the results were mixed, with both negative and positive but in most cases statistically insignificant effects. Furthermore, our results suggested that this heterogeneity in belief effects is important for explaining country differences in redistribution support. With a few exceptions (Germany, Brazil and to some extent Russia), a quite substantial share of countries' stronger support for redistribution relative to the US could be explained by people in these countries seemingly assigning a lower degree of responsibility to inputs believed to be important for income determination. So, while differences in beliefs on what causes income differences seem to be important for explaining within-country variation in redistributive preferences, they do little to explain between-country differences. Differences in the *effects* of holding certain beliefs, however, appear important for explaining between-country variation in redistributive preferences.

Summing up, our findings indicate that self-interest considerations as well as input-based fairness concerns contribute to explain redistributive preferences, but also that there is substantial country variation, in redistribution support, in key factors explaining redistribution support, as well as in the *effects* of these factors *on* redistribution support. In particular, looking at our results the country most studied in this field – the US – is quite an extreme case, displaying the lowest support for redistribution, the most positive views about the rewards to effort and intelligence/skills, and some of the strongest effects of our belief variables. This tells us that in trying to understand fairness-based and self-interested motivations behind preferences for redistribution we cannot focus on one-country alone – we need to evaluate both within and between-country variations.

Acknowledgements

We would like to thank two anonymous referees, Olof Johansson-Stenman, Arne Bigsten, Karl Ove Moene, colleagues at University of Gothenburg, and seminar participants at the First Nordic Workshop on Behavioural and Experimental Economics in Oslo 2006, the workshop Equality: Theory, Political Economy, and Prospects in Bergen 2007, and at the 2007 EEA conference in Budapest for valuable suggestions and discussion.

Appendix A.

See Tables A1–A6.

Table A1
Variable description.

Variable	Description
Preferences for redistribution	The response to the statement, '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*'; 1 if respondent chooses strongly disagree, 2 if respondent chooses disagree, 3 if respondent chooses neither agree nor disagree, 4 if respondent chooses agree and 5 if respondent chooses strongly agree.
Belief variables	
Believe wealthy family important	1 if respondent answers *essential* or *very important* to the question, '*For getting ahead, how important is coming from a wealthy family*?'; 0 otherwise.
No strong family belief	1 if respondent answers *fairly important* to the question, '*For getting ahead, how important is coming from a wealthy family*?'; 0 otherwise.
Believe family not important	1 if respondent answers *not very important* or *not important at all* to the question, '*For getting ahead, how important is coming from a wealthy family*?'; 0 otherwise. Used as reference category in estimation.
Believe intelligence and skills rewarded	1 if respondent responds *agree* or *strongly agree* to the statement, '*In [country] people get rewarded for their intelligence and skills*'; 0 otherwise.
No strong intelligence/skills belief	1 if respondent responds *neither agree nor disagree* to the statement, '*In [country] people get rewarded for their intelligence and skills*'; 0 otherwise.
Believe intelligence and skills not rewarded	1 if respondent responds *disagree* or *strongly disagree* to the statement, '*In [country] people get rewarded for their intelligence and skills*'; 0 otherwise. Used as reference category in estimation.
Believe effort rewarded	1 if respondent responds *agree* or *strongly agree* to the statement, '*In [country] people get rewarded for their effort*'; 0 otherwise.
No strong effort belief	1 if respondent responds *neither agree nor disagree* to the statement, '*In [country] people get rewarded for their effort*'; 0 otherwise.
Believe effort not rewarded	1 if respondent responds *disagree* or *strongly disagree* to the statement, '*In [country] people get rewarded for their effort*'; 0 otherwise. Used as reference category in estimation.
Self-interest variable	
Relative income	Household income per adult equivalent divided by the country sample average. A common country average was used for eastern and western Germany.
Control variables	
Age	Age in years
Female	1 if female; 0 else
Higher education	1 if respondent has some post secondary school education; 0 else
Father has higher education	1 if respondent's father has completed secondary school; 0 else

Table A1 (*Continued*)

Variable	Description
Upper class	1 if respondent's self reported class is *upper class* or *upper middle class*; 0 else
Working class	1 if respondent's self reported class is *working class* or *lower class*; 0 else
Middle class	1 if respondent's self reported class is *middle class*; 0 else. Used as reference category in estimation.
Inequality necessary for prosperity	1 if respondent respond *agree* or *strongly agree* to the statement, '*Large differences in income are necessary for [country's] prosperity*'; 0 *otherwise*
Country dummies	1 if respondent belongs to the country in question; 0 else. USA used as reference category in estimations.

Table A2
Log-likelihood ratio tests.

Restricted model	Unrestricted model	LR χ^2	*p*-Value
Panel 1: Joint importance of belief variables			
Pooled, excluding belief vars.	Pooled sample benchmark	295.80	0.000
Australia, excluding belief vars.	Australia benchmark	36.99	0.000
Brazil, excluding belief vars.	Brazil benchmark	17.54	0.008
Canada, excluding belief vars.	Canada benchmark	22.39	0.001
Chile, excluding belief vars.	Chile benchmark	8.97	0.175
Cyprus, excluding belief vars.	Cyprus benchmark	51.27	0.000
Czech Rep., excluding belief vars.	Czech Rep. benchmark	28.52	0.000
Denmark, excluding belief vars.	Denmark benchmark	21.27	0.002
France, excluding belief vars.	France benchmark	15.84	0.015
Germany (west), excluding belief vars.	Germany (west) benchmark	15.21	0.019
Germany (east), excluding belief vars.	Germany (east) benchmark	16.92	0.010
Hungary, excluding belief vars.	Hungary benchmark	15.96	0.014
Ireland, excluding belief vars.	Ireland benchmark	11.97	0.063
Japan, excluding belief vars.	Japan benchmark	12.37	0.054
Latvia, excluding belief vars.	Latvia benchmark	19.10	0.004
New Zealand, excluding belief vars.	New Zealand benchmark	20.86	0.002
Norway, excluding belief vars.	Norway benchmark	20.21	0.003
Philippines, excluding belief vars.	Philippines benchmark	26.03	0.000
Poland, excluding belief vars.	Poland benchmark	35.35	0.000
Portugal, excluding belief vars.	Portugal benchmark	21.00	0.002
Russia, excluding belief vars.	Russia benchmark	21.14	0.002
Slovenia, excluding belief vars.	Slovenia	7.43	0.283
Spain, excluding belief vars.	Spain benchmark	20.21	0.003
Sweden, excluding belief vars.	Sweden benchmark	17.73	0.007
Switzerland, excluding belief vars.	Switzerland benchmark	9.44	0.150
USA, excluding belief vars.	USA benchmark	38.95	0.000
Panel 2: Tests of parameter homogeneity			
Pooled sample benchmark	Allow belief parameters to vary for each belief and each country	247.03	0.000
Pooled sample benchmark	Allow family belief parameters to vary for each country	76.78	0.005
Pooled sample benchmark	Allow effort belief parameters to vary for each country	108.77	0.000
Pooled sample benchmark	Allow intelligence and skills belief parameters to vary for each country	98.55	0.000

Based on ordered probit estimations where the dependent variable is the answers to the statement, '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*', ranging from 1 for *strongly disagree* to 5 for *strongly agree*. The belief variables are responses to whether effort and intelligence/skills are rewarded, and to whether it is important to be from a wealthy family to get ahead. These and other explanatory variables in the benchmark model are described in Table A1.

Table A3
One-sided tests of coefficients from ordered probit estimation[a] of the probability to agree with the redistributive statement[b].

Sample	*p*-Value of testing the null hypothesis				
	FAM[c] ≤ 0	EFF[d] ≥ 0	SKI[e] $\geq$ FAM	EFF $\geq$ SKI	EFF $\geq$ FAM
Pooled	0.000	0.000	0.000	0.063	0.000
Australia	0.000	0.019	0.001	0.182	0.000
Brazil	0.098	0.880	0.014	0.957	0.507
Canada	0.270	0.064	0.011	0.704	0.055
Chile	0.124	0.200	0.126	0.457	0.080
Cyprus	0.035	0.036	0.001	0.646	0.005
Czech rep.	0.386	0.011	0.086	0.309	0.023

Table A3 (*Continued*)

Sample	*p*-Value of testing the null hypothesis				
	FAM[c] ≤ 0	EFF[d] ≥ 0	SKI[e] ≥ FAM	EFF ≥ SKI	EFF ≥ FAM
Denmark	0.000	0.385	0.131	0.045	0.001
France	0.009	0.070	0.032	0.207	0.003
Germany (west)	0.176	0.173	0.094	0.504	0.091
Germany (east)	0.012	0.359	0.069	0.470	0.048
Hungary	0.023	0.054	0.122	0.112	0.008
Ireland	0.030	0.243	0.004	0.772	0.075
Japan	0.013	0.097	0.009	0.495	0.005
Latvia	0.002	0.681	0.000	0.936	0.085
New Zealand	0.012	0.020	0.013	0.262	0.000
Norway	0.000	0.658	0.000	0.768	0.001
Philippines	0.000	0.837	0.004	0.830	0.057
Poland	0.000	0.452	0.006	0.245	0.000
Portugal	0.008	0.016	0.175	0.043	0.001
Russia	0.001	0.224	0.033	0.389	0.017
Slovenia	0.450	0.269	0.270	0.490	0.280
Spain	0.145	0.001	0.391	0.017	0.001
Sweden	0.001	0.541	0.013	0.590	0.024
Switzerland	0.038	0.223	0.021	0.589	0.052
USA	0.000	0.006	0.005	0.068	0.000

[a] Based on estimation of Eq. (1).
[b] '*It is the responsibility of the government to reduce the difference in income between people with high incomes and those with low incomes*'.
[c] Coefficient of 'Believe wealthy family important'.
[d] Coefficient of 'Believe effort rewarded'.
[e] Coefficient of 'Believe intelligence and skills rewarded'.

Table A4
Relative income effects[a] on the probability of different responses to the question '*For getting ahead, how important is coming from a wealthy family*?'.

Sample	Not at all important	Not very important	Fairly important	Very important	Essential
Pooled (25 countries)	0.001***	0.001***	0.000***	−0.002***	−0.001***
Pooled (20 countries)[b]	0.000	0.001	0.000	−0.001	−0.000
Australia	0.003	0.004	−0.002	−0.004	−0.001
Brazil	0.001	0.000	0.000	0.000	−0.001
Canada	−0.009	−0.005	0.006	0.006	0.002
Chile	−0.000	−0.000	0.000	0.000	0.000
Cyprus	−0.000	−0.000	−0.001	0.000	0.001
Czech rep.	−0.001	−0.001	0.001	0.001	0.001
Denmark	0.003	0.002	−0.003	−0.002	−0.001
France	−0.003	−0.002	0.002	0.002	0.000
Germany (west)	0.000	0.000	−0.000	−0.000	−0.000
Germany (east)	−0.004	−0.006	−0.000	0.007	0.003
Hungary	0.004**	0.006**	0.000	−0.004**	−0.006**
Ireland	0.002	0.003	−0.001	−0.004	−0.001
Japan	−0.001	−0.001	0.001	0.001	0.000
Latvia	0.004**	0.003**	0.001*	−0.005**	−0.004**
New Zealand	−0.000	−0.000	0.000	0.000	0.000
Norway	0.002	0.003	−0.003	−0.002	−0.000
Philippines	0.001	0.001	0.000	−0.001	−0.001
Poland	0.001	0.002	0.004	−0.002	−0.005
Portugal	−0.003	−0.005	−0.002	0.003	0.007
Russia	0.003*	0.003*	0.001*	−0.002*	−0.004*
Slovenia	−0.004	−0.004	0.001	0.004	0.002
Spain	0.002	0.003	0.002	−0.005	−0.003
Sweden	0.002	0.003	−0.002	−0.003	−0.001
Switzerland	−0.006	−0.007	0.005	0.005	0.003
USA	0.007	0.005	−0.004	−0.007	−0.002

The effects are from ordered probit estimations where the dependent variable is the answer to the question '*For getting ahead, how important is coming from a wealthy family*?', and the explanatory variables included are: relative income, age, female, higher education, father has higher education, upper class, working class, inequality necessary for prosperity, and country dummies.
[a] Measures the effects of a relative income increase of one median absolute deviation increase around the median.
[b] Brazil, France, Hungary, Latvia, and Russia excluded.
* Significance at the 10% level.
** Significance at the 5% level.
*** Significance at the 1% level.

Table A5
Relative income effects[a] on the probability of agreeing with the statement, '*In [country] people get rewarded for their effort*'.

Sample	Strongly disagree	Disagree	Neither	Agree	Strongly agree
Pooled (25 countries)	0.001**	0.001**	0.000**	−0.001**	−0.001**
Pooled (20 countries)[b]	0.000	0.000	0.000	−0.000	−0.000
Australia	−0.000	−0.003	−0.002	0.004	0.001
Brazil	0.011***	0.001***	−0.001***	−0.003***	−0.008***
Canada	−0.001	−0.004	−0.002	0.006	0.001
Chile	0.001	0.002	0.000	−0.002	−0.001
Cyprus	0.002	0.003	−0.001	−0.003	−0.001
Czech rep.	−0.001	0.000	0.000	0.000	0.000
Denmark	−0.001	−0.001	0.000	0.001	0.001
France	−0.008***	−0.011***	0.005***	0.012***	0.002**
Germany (west)	−0.000	−0.002	−0.002	0.004	0.000
Germany (east)	−0.001	−0.002	−0.001	0.003	0.000
Hungary	−0.010**	0.000	0.005**	0.003**	0.002**
Ireland	0.001	0.005	0.001	−0.006	−0.001
Japan	0.003	0.002	0.001	−0.004	−0.003
Latvia	0.002	0.000	−0.001	−0.001	−0.000
New Zealand	−0.002	−0.008	−0.001	0.008	0.003
Norway	0.002	0.004	−0.001	−0.005	−0.001
Philippines	0.000	0.000	0.000	−0.000	−0.000
Poland	−0.002	−0.003	0.001	0.003	0.001
Portugal	−0.002	−0.001	0.000	0.002	0.001
Russia	−0.007*	0.003*	0.002*	0.001*	0.001*
Slovenia	0.006	0.003	−0.004	−0.003	−0.001
Spain	−0.008**	−0.010**	0.001	0.014**	0.003**
Sweden	−0.001	−0.002	−0.001	0.003	0.001
Switzerland	−0.001	−0.002	−0.002	0.003	0.001
USA	−0.001	−0.005	−0.005	0.006	0.005

The effects are from ordered probit estimations where the dependent variable is the answer to the question '*In [country] people get rewarded for their effort*', and the explanatory variables included are: relative income, age, female, higher education, father has higher education, upper class, working class, inequality necessary for prosperity, and country dummies.

[a] Measures the effects of a relative income increase of one median absolute deviation increase around the median.
[b] Brazil, France, Hungary, Latvia, and Russia are excluded.
* Significance at the 10% level.
** Significance at the 5% level.
*** Significance at the 1% level.

Table A6
Relative income effects[a] on the probability of agreeing with the statement, '*In [country] people get rewarded for their intelligence and skills*'.

Estimation sample	Strongly disagree	Disagree	Neither	Agree	Strongly agree
Pooled (25 countries)	0.000	0.000	0.000	−0.000	−0.000
Pooled (20 countries)[b]	0.000	0.001	0.000	−0.001	−0.000
Australia	0.000	0.003	0.002	−0.004	−0.002
Brazil	0.003*	0.001*	0.000	−0.001*	−0.003*
Canada	−0.002	−0.008	−0.005	0.011	0.004
Chile	0.001	0.001	0.000	−0.001	−0.001
Cyprus	0.001	0.001	−0.000	−0.001	−0.000
Czech rep.	−0.001	0.000	0.000	0.000	0.000
Denmark	0.002	0.003	0.003	−0.003	−0.005
France	−0.004*	−0.009*	0.000	0.011**	0.002*
Germany (west)	−0.000	−0.001	−0.000	0.001	0.000
Germany (east)	0.000	0.002	0.002	−0.003	−0.001
Hungary	−0.001	−0.001	0.000	0.001	0.000
Ireland	0.002*	0.013*	0.006*	−0.015*	−0.006*
Japan	0.002	0.002	0.003	−0.003	−0.004
Latvia	−0.007**	−0.002**	0.003**	0.005**	0.001**
New Zealand	0.000	0.001	0.001	−0.002	−0.001
Norway	0.003**	0.009**	0.001*	−0.012**	−0.002**
Philippines	−0.001*	−0.002*	−0.002*	0.001*	0.004*
Poland	−0.002	−0.003	0.001	0.003	0.001
Portugal	0.006	0.004	0.000	−0.006	−0.004
Russia	−0.007**	0.002**	0.002**	0.002**	0.001*
Slovenia	0.001	0.001	−0.000	−0.001	−0.000
Spain	−0.000	−0.000	−0.000	0.000	0.000

Table A6 (*Continued*)

Estimation sample	Strongly disagree	Disagree	Neither	Agree	Strongly agree
Sweden	0.000	0.000	0.000	−0.000	−0.000
Switzerland	−0.001	−0.004	−0.008	0.010	0.003
USA	−0.001	−0.005	−0.006	0.004	0.008

The effects are from ordered probit estimations where the dependent variable is the answer to the question '*In [country] people get rewarded for their effort*', and the explanatory variables included are: relative income, age, female, higher education, father has higher education, upper class, working class, inequality necessary for prosperity, and country dummies.

[a] Measures the effects of a relative income increase of one median absolute deviation increase around the median.

[b] Brazil, France, Hungary, Latvia, and Russia are excluded.

[*] Significance at the 10% level.

[**] Significance at the 5% level.

References

Adams, J.S., 1965. Inequity in social exchange. In: Berkowitz, L. (Ed.), Advances in Experimental Social Psychology, vol. 2. Academic Press, New York, pp. 267–299.
Alesina, A., Angeletos, G.-M., 2005. Fairness and redistribution. American Economic Review 95, 960–980.
Alesina, A., Fuchs-Schündeln, N., 2007. Goodbye Lenin (or not?): the effect of communism on people. American Economic Review 97, 1507–1528.
Alesina, A., Glaeser, E., Sacerdote, B., 2001. Why doesn't the United States have a European-style welfare state? Brookings Papers on Economic Activity 2001, 187–254.
Alesina, A., La Ferrara, E., 2005. Preferences for redistribution in the land of opportunities. Journal of Public Economics 89, 897–931.
Bénabou, R., Ok, E., 2001. Social mobility and the demand for redistribution: the POUM hypothesis. The Quarterly Journal of Economics 116, 447–487.
Buchanan, J.M., Tullock, G., 1962. The Calculus of Consent: Logical Foundations of Constitutional Democracy. University of Michigan Press, Ann Arbor.
Burrows, P., Loomes, G., 1994. The impact of fairness on bargaining behaviour. Empirical Economics 19, 201–221.
Cappelen, A.W., Hole, A.D., Sørensen, E.Ø., Tungodden, B., 2007. The pluralism of fairness ideals: an experimental approach. The American Economic Review 97, 818–827.
Cappelen, A.W., Tungodden, B., 2006. Relocating the responsibility cut: should more responsibility imply less redistribution? Politics, Philosophy and Economics 5, 353–362.
Clark, J., 1998. Fairness in public good provision: an investigation of preferences for equality and proportionality. Canadian Journal of Economics 31, 708–728.
Conover, W.J., 1999. Practical Nonparametric Statistics. John Wiley and Sons Inc., New York.
Corneo, G., Grüner, H.P., 2000. Social limits to redistribution. American Economic Review 90, 1491–1507.
Corneo, G., Grüner, H.P., 2002. Individual preferences for political redistribution. Journal of Public Economics 83, 83–107.
Dworkin, R., 1981a. What is equality? Part 1. Equality of welfare. Philosophy and Public Affairs 10, 185–246.
Dworkin, R., 1981b. What is equality? Part 2. Equality of resources. Philosophy and Public Affairs 10, 283–345.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition and cooperation. Quarterly Journal of Economics 114, 817–868.
Fong, C., 2001. Social preferences, self-interest, and the demand for redistribution. Journal of Public Economics 82, 225–246.
Fong, C., Bowles, S., Gintis, H., 2005. Behavioural motives for income redistribution. The Australian Economic Review 38, 285–297.
Goodman, L.A., 1954. Kolmogorov–Smirnov tests for psychological research. Psychological Bulletin 2, 160–168.
Konow, J., 2003. Which is the fairest one of all? A positive analysis of justice theories. Journal of Economic Literature 41, 1188–1239.
Noether, G.E., 1963. Note on the Kolmogorov statistic in the discrete case. Metrika 7, 115–116.
Piketty, T., 1995. Social mobility and redistributive politics. The Quarterly Journal of Economics 110, 551–584.
Ravallion, M., Lokshin, M., 2000. Who wants to redistribute? The tunnel effect in 1990s Russia. Journal of Public Economics 76, 87–104.
Roemer, J.E., 2002. Equality of opportunity: a progress report. Social Choice and Welfare 19, 455–471.
Van Dijk, E., Wilke, H., 1994. Asymmetry of wealth and public good provision. Social Psychology Quarterly 57, 352–359.

Previous doctoral theses in the Department of Economics, Gothenburg

Avhandlingar publicerade innan serien Ekonomiska Studier startades
(Theses published before the series Ekonomiska Studier was started):

Östman, Hugo (1911), Norrlands ekonomiska utveckling
Moritz, Marcus (1911), Den svenska tobaksindustrien
Sundbom, I. (1933), Prisbildning och ändamålsenlighet
Gerhard, I. (1948), Problem rörande Sveriges utrikeshandel 1936/38
Hegeland, Hugo (1951), The Quantity Theory of Money
Mattsson, Bengt (1970), Cost-Benefit analys
Rosengren, Björn (1975), Valutareglering och nationell ekonomisk politik
Hjalmarsson, Lennart (1975), Studies in a Dynamic Theory of Production and its Applications
Örtendahl, Per-Anders (1975), Substitutionsaspekter på produktionsprocessen vid massaframställning
Anderson, Arne M. (1976), Produktion, kapacitet och kostnader vid ett helautomatiskt emballageglasbruk
Ohlsson, Olle (1976), Substitution och odelbarheter i produktionsprocessen vid massaframställning
Gunnarsson, Jan (1976), Produktionssystem och tätortshierarki – om sambandet mellan rumslig och ekonomisk struktur
Köstner, Evert (1976), Optimal allokering av tid mellan utbildning och arbete
Wigren, Rune (1976), Analys av regionala effektivitetsskillnader inom industribranscher
Wästlund, Jan (1976), Skattning och analys av regionala effektivitetsskillnader inom industribranscher
Flöjstad, Gunnar (1976), Studies in Distortions, Trade and Allocation Problems
Sandelin, Bo (1977), Prisutveckling och kapitalvinster på bostadsfastigheter
Dahlberg, Lars (1977), Empirical Studies in Public Planning
Lönnroth, Johan (1977), Marxism som matematisk ekonomi
Johansson, Börje (1978), Contributions to Sequential Analysis of Oligopolistic Competition

Ekonomiska Studier, utgivna av Nationalekonomiska institutionen vid Göteborgs Universitet. Nr 1 och 4 var inte doktorsavhandlingar. (The contributions to the department series 'Ekonomiska Studier' where no. 1 and 4 were no doctoral theses):

2. **Ambjörn, Erik** (1959), Svenskt importberoende 1926-1956: en ekonomisk-statistisk kartläggning med kommentarer
3. **Landgren, K-G.** (1960), Den "Nya ekonomien" i Sverige: J.M. Keynes, E. Wigfors och utecklingen 1927-39
5. **Bigsten, Arne** (1979), Regional Inequality and Development: A Case Study of Kenya
6. **Andersson, Lars** (1979), Statens styrning av de kommunala budgetarnas struktur (Central Government Influence on the Structure of the Municipal Budget)
7. **Gustafsson, Björn** (1979), Inkomst- och uppväxtförhållanden (Income and Family Background)

8. **Granholm, Arne** (1981), Interregional Planning Models for the Allocation of Private and Public Investments
9. **Lundborg, Per** (1982), Trade Policy and Development: Income Distributional Effects in the Less Developed Countries of the US and EEC Policies for Agricultural Commodities
10. **Juås, Birgitta** (1982), Värdering av risken för personskador. En jämförande studie av implicita och explicita värden. (Valuation of Personal Injuries. A comparison of Explicit and Implicit Values)
11. **Bergendahl, Per-Anders** (1982), Energi och ekonomi - tillämpningar av input-output analys (Energy and the Economy - Applications of Input-Output Analysis)
12. **Blomström, Magnus** (1983), Foreign Investment, Technical Efficiency and Structural Change - Evidence from the Mexican Manufacturing Industry
13. **Larsson, Lars-Göran** (1983), Comparative Statics on the Basis of Optimization Methods
14. **Persson, Håkan** (1983), Theory and Applications of Multisectoral Growth Models
15. **Sterner, Thomas** (1986), Energy Use in Mexican Industry.
16. **Flood, Lennart** (1986), On the Application of Time Use and Expenditure Allocation Models.
17. **Schuller, Bernd-Joachim** (1986), Ekonomi och kriminalitet - en empirisk undersökning av brottsligheten i Sverige (Economics of crime - an empirical analysis of crime in Sweden)
18. **Walfridson, Bo** (1987), Dynamic Models of Factor Demand. An Application to Swedish Industry.
19. **Stålhammar, Nils-Olov** (1987), Strukturomvandling, företagsbeteende och förväntningsbildning inom den svenska tillverkningsindustrin (Structural Change, Firm Behaviour and Expectation Formation in Swedish Manufactury)
20. **Anxo, Dominique** (1988), Sysselsättningseffekter av en allmän arbetstidsförkortning (Employment effects of a general shortage of the working time)
21. **Mbelle, Ammon** (1988), Foreign Exchange and Industrial Development: A Study of Tanzania.
22. **Ongaro, Wilfred** (1988), Adoption of New Farming Technology: A Case Study of Maize Production in Western Kenya.
23. **Zejan, Mario** (1988), Studies in the Behavior of Swedish Multinationals.
24. **Görling, Anders** (1988), Ekonomisk tillväxt och miljö. Förorenings-struktur och ekonomiska effekter av olika miljövårdsprogram. (Economic Growth and Environment. Pollution Structure and Economic Effects of Some Environmental Programs).
25. **Aguilar, Renato** (1988), Efficiency in Production: Theory and an Application on Kenyan Smallholders.
26. **Kayizzi-Mugerwa, Steve** (1988), External Shocks and Adjustment in Zambia.

27. **Bornmalm-Jardelöw, Gunilla** (1988), Högre utbildning och arbetsmarknad (Higher Education and the Labour Market)
28. **Tansini, Ruben** (1989), Technology Transfer: Dairy Industries in Sweden and Uruguay.
29. **Andersson, Irene** (1989), Familjebeskattning, konsumtion och arbetsutbud - En ekonometrisk analys av löne- och inkomstelasticiteter samt policysimuleringar för svenska hushåll (Family Taxation, Consumption and Labour Supply - An Econometric Analysis of Wage and Income Elasticities and Policy Simulations for Swedish Households)
30. **Henrekson, Magnus** (1990), An Economic Analysis of Swedish Government Expenditure
31. **Sjöö, Boo** (1990), Monetary Policy in a Continuous Time Dynamic Model for Sweden
32. **Rosén, Åsa** (1991), Contributions to the Theory of Labour Contracts.
33. **Loureiro, Joao M. de Matos** (1992), Foreign Exchange Intervention, Sterilization and Credibility in the EMS: An Empirical Study
34. **Irandoust, Manuchehr** (1993), Essays on the Behavior and Performance of the Car Industry
35. **Tasiran, Ali Cevat** (1993), Wage and Income Effects on the Timing and Spacing of Births in Sweden and the United States
36. **Milopoulos, Christos** (1993), Investment Behaviour under Uncertainty: An Econometric Analysis of Swedish Panel Data
37. **Andersson, Per-Åke** (1993), Labour Market Structure in a Controlled Economy: The Case of Zambia
38. **Storrie, Donald W.** (1993), The Anatomy of a Large Swedish Plant Closure
39. **Semboja, Haji Hatibu Haji** (1993), Energy and Development in Kenya
40. **Makonnen, Negatu** (1993), Labor Supply and the Distribution of Economic Well-Being: A Case Study of Lesotho
41. **Julin, Eva** (1993), Structural Change in Rural Kenya
42. **Durevall, Dick** (1993), Essays on Chronic Inflation: The Brazilian Experience
43. **Veiderpass, Ann** (1993), Swedish Retail Electricity Distribution: A Non-Parametric Approach to Efficiency and Productivity Change
44. **Odeck, James** (1993), Measuring Productivity Growth and Efficiency with Data Envelopment Analysis: An Application on the Norwegian Road Sector
45. **Mwenda, Abraham** (1993), Credit Rationing and Investment Behaviour under Market Imperfections: Evidence from Commercial Agriculture in Zambia
46. **Mlambo, Kupukile** (1993), Total Factor Productivity Growth: An Empirical Analysis of Zimbabwe's Manufacturing Sector Based on Factor Demand Modelling
47. **Ndung'u, Njuguna** (1993), Dynamics of the Inflationary Process in Kenya
48. **Modén, Karl-Markus** (1993), Tax Incentives of Corporate Mergers and Foreign Direct Investments
49. **Franzén, Mikael** (1994), Gasoline Demand - A Comparison of Models
50. **Heshmati, Almas** (1994), Estimating Technical Efficiency, Productivity Growth And Selectivity Bias Using Rotating Panel Data: An Application to Swedish Agriculture

51. **Salas, Osvaldo** (1994), Efficiency and Productivity Change: A Micro Data Case Study of the Colombian Cement Industry
52. **Bjurek, Hans** (1994), Essays on Efficiency and Productivity Change with Applications to Public Service Production
53. **Cabezas Vega, Luis** (1994), Factor Substitution, Capacity Utilization and Total Factor Productivity Growth in the Peruvian Manufacturing Industry
54. **Katz, Katarina** (1994), Gender Differentiation and Discrimination. A Study of Soviet Wages
55. **Asal, Maher** (1995), Real Exchange Rate Determination and the Adjustment Process: An Empirical Study in the Cases of Sweden and Egypt
56. **Kjulin, Urban** (1995), Economic Perspectives on Child Care
57. **Andersson, Göran** (1995), Volatility Forecasting and Efficiency of the Swedish Call Options Market
58. **Forteza, Alvaro** (1996), Credibility, Inflation and Incentive Distortions in the Welfare State
59. **Locking, Håkan** (1996), Essays on Swedish Wage Formation
60. **Välilä, Timo** (1996), Essays on the Credibility of Central Bank Independence
61. **Yilma, Mulugeta** (1996), Measuring Smallholder Efficiency: Ugandan Coffee and Food-Crop Production
62. **Mabugu, Ramos E.** (1996), Tax Policy Analysis in Zimbabwe Applying General Equilibrium Models
63. **Johansson, Olof** (1996), Welfare, Externalities, and Taxation; Theory and Some Road Transport Applications.
64. **Chitiga, Margaret** (1996), Computable General Equilibrium Analysis of Income Distribution Policies in Zimbabwe
65. **Leander, Per** (1996), Foreign Exchange Market Behavior Expectations and Chaos
66. **Hansen, Jörgen** (1997), Essays on Earnings and Labor Supply
67. **Cotfas, Mihai** (1997), Essays on Productivity and Efficiency in the Romanian Cement Industry
68. **Horgby, Per-Johan** (1997), Essays on Sharing, Management and Evaluation of Health Risks
69. **Nafar, Nosratollah** (1997), Efficiency and Productivity in Iranian Manufacturing Industries
70. **Zheng, Jinghai** (1997), Essays on Industrial Structure, Technical Change, Employment Adjustment, and Technical Efficiency
71. **Isaksson, Anders** (1997), Essays on Financial Liberalisation in Developing Countries: Capital mobility, price stability, and savings
72. **Gerdin, Anders** (1997), On Productivity and Growth in Kenya, 1964-94
73. **Sharifi, Alimorad** (1998), The Electricity Supply Industry in Iran: Organization, performance and future development
74. **Zamanian, Max** (1997), Methods for Mutual Fund Portfolio Evaluation: An application to the Swedish market
75. **Manda, Damiano Kulundu** (1997), Labour Supply, Returns to Education, and the Effect of Firm Size on Wages: The case of Kenya
76. **Holmén, Martin** (1998), Essays on Corporate Acquisitions and Stock Market Introductions

77. **Pan, Kelvin** (1998), Essays on Enforcement in Money and Banking
78. **Rogat, Jorge** (1998), The Value of Improved Air Quality in Santiago de Chile
79. **Peterson, Stefan** (1998), Essays on Large Shareholders and Corporate Control
80. **Belhaj, Mohammed** (1998), Energy, Transportation and Urban Environment in Africa: The Case of Rabat-Salé, Morocco
81. **Mekonnen, Alemu** (1998), Rural Energy and Afforestation: Case Studies from Ethiopia
82. **Johansson, Anders** (1998), Empirical Essays on Financial and Real Investment Behavior
83. **Köhlin, Gunnar** (1998), The Value of Social Forestry in Orissa, India
84. **Levin, Jörgen** (1998), Structural Adjustment and Poverty: The Case of Kenya
85. **Ncube, Mkhululi** (1998), Analysis of Employment Behaviour in Zimbabwe
86. **Mwansa, Ladslous** (1998), Determinants of Inflation in Zambia
87. **Agnarsson, Sveinn** (1998), Of Men and Machines: Essays in Applied Labour and Production Economics
88. **Kadenge, Phineas** (1998), Essays on Macroeconomic Adjustment in Zimbabwe: Inflation, Money Demand, and the Real Exchange Rate
89. **Nyman, Håkan** (1998), An Economic Analysis of Lone Motherhood in Sweden
90. **Carlsson, Fredrik** (1999), Essays on Externalities and Transport
91. **Johansson, Mats** (1999), Empirical Studies of Income Distribution
92. **Alemu, Tekie** (1999), Land Tenure and Soil Conservation: Evidence from Ethiopia
93. **Lundvall, Karl** (1999), Essays on Manufacturing Production in a Developing Economy: Kenya 1992-94
94. **Zhang, Jianhua** (1999), Essays on Emerging Market Finance
95. **Mlima, Aziz Ponary** (1999), Four Essays on Efficiency and Productivity in Swedish Banking
96. **Davidsen, Björn-Ivar** (2000), Bidrag til den økonomisk-metodologiske tenkningen (Contributions to the Economic Methodological Thinking)
97. **Ericson, Peter** (2000), Essays on Labor Supply
98. **Söderbom, Måns** (2000), Investment in African Manufacturing: A Microeconomic Analysis
99. **Höglund, Lena** (2000), Essays on Environmental Regulation with Applications to Sweden
100. **Olsson, Ola** (2000), Perspectives on Knowledge and Growth
101. **Meuller, Lars** (2000), Essays on Money and Credit
102. **Österberg, Torun** (2000), Economic Perspectives on Immigrants and Intergenerational Transmissions
103. **Kalinda Mkenda, Beatrice** (2001), Essays on Purchasing Power Parity, RealExchange Rate, and Optimum Currency Areas
104. **Nerhagen, Lena** (2001), Travel Demand and Value of Time - Towards an Understanding of Individuals Choice Behavior

105. **Mkenda, Adolf** (2001), Fishery Resources and Welfare in Rural Zanzibar
106. **Eggert, Håkan** (2001), Essays on Fisheries Economics
107. **Andrén, Daniela** (2001), Work, Sickness, Earnings, and Early Exits from the Labor Market. An Empirical Analysis Using Swedish Longitudinal Data
108. **Nivorozhkin, Eugene** (2001), Essays on Capital Structure
109. **Hammar, Henrik** (2001), Essays on Policy Instruments: Applications to Smoking and the Environment
110. **Nannyonjo, Justine** (2002), Financial Sector Reforms in Uganda (1990-2000): Interest Rate Spreads, Market Structure, Bank Performance and Monetary Policy
111. **Wu, Hong** (2002), Essays on Insurance Economics
112. **Linde-Rahr, Martin** (2002), Household Economics of Agriculture and Forestry in Rural Vienam
113. **Maneschiöld, Per-Ola** (2002), Essays on Exchange Rates and Central Bank Credibility
114. **Andrén, Thomas** (2002), Essays on Training, Welfare and Labor Supply
115. **Granér, Mats** (2002), Essays on Trade and Productivity: Case Studies of Manufacturing in Chile and Kenya
116. **Jaldell, Henrik** (2002), Essays on the Performance of Fire and Rescue Services
117. **Alpizar, Francisco, R.** (2002), Essays on Environmental Policy-Making in Developing Countries: Applications to Costa Rica
118. **Wahlberg, Roger** (2002), Essays on Discrimination, Welfare and Labor Supply
119. **Piculescu, Violeta** (2002), Studies on the Post-Communist Transition
120. **Pylkkänen, Elina** (2003), Studies on Household Labor Supply and Home Production
121. **Löfgren, Åsa** (2003), Environmental Taxation – Empirical and Theoretical Applications
122. **Ivaschenko, Oleksiy** (2003), Essays on Poverty, Income Inequality and Health in Transition Economies
123. **Lundström, Susanna** (2003), On Institutions, Economic Growth and the Environment
124. **Wambugu, Anthony** (2003), Essays on Earnings and Human Capital in Kenya
125. **Adler, Johan** (2003), Aspects of Macroeconomic Saving
126. **Erlandsson, Mattias** (2003), On Monetary Integration and Macroeconomic Policy
127. **Brink, Anna** (2003), On the Political Economy of Municipality Break-Ups
128. **Ljungwall, Christer** (2003), Essays on China's Economic Performance During the Reform Period
129. **Chifamba, Ronald** (2003), Analysis of Mining Investments in Zimbabwe
130. **Muchapondwa, Edwin** (2003), The Economics of Community-Based Wildlife Conservation in Zimbabwe
131. **Hammes, Klaus** (2003), Essays on Capital Structure and Trade Financing
132. **Abou-Ali, Hala** (2003), Water and Health in Egypt: An Empirical Analysis
133. **Simatele, Munacinga** (2004), Financial Sector Reforms and Monetary Policy in Zambia
134. **Tezic, Kerem** (2004), Essays on Immigrants' Economic Integration
135. INSTÄLLD

136. **Gjirja, Matilda** (2004), Efficiency and Productivity in Swedish Banking
137. **Andersson, Jessica** (2004), Welfare Environment and Tourism in Developing Countries
138. **Chen, Yinghong** (2004), Essays on Voting Power, Corporate Governance and Capital Structure
139. **Yesuf, Mahmud** (2004), Risk, Time and Land Management under Market Imperfections: Applications to Ethiopia
140. **Kateregga, Eseza** (2005), Essays on the Infestation of Lake Victoria by the Water Hyacinth
141. **Edvardsen, Dag Fjeld** (2004), Four Essays on the Measurement of Productive Efficiency
142. **Lidén, Erik** (2005), Essays on Information and Conflicts of Interest in Stock Recommendations
143. **Dieden, Sten** (2005), Income Generation in the African and Coloured Population – Three Essays on the Origins of Household Incomes in South Africa
144. **Eliasson, Marcus** (2005), Individual and Family Consequences of Involuntary Job Loss
145. **Mahmud, Minhaj** (2005), Measuring Trust and the Value of Statistical Lives: Evidence from Bangladesh
146. **Lokina, Razack Bakari** (2005), Efficiency, Risk and Regulation Compliance: Applications to Lake Victoria Fisheries in Tanzania
147. **Jussila Hammes, Johanna** (2005), Essays on the Political Economy of Land Use Change
148. **Nyangena, Wilfred** (2006), Essays on Soil Conservation, Social Capital and Technology Adoption
149. **Nivorozhkin, Anton** (2006), Essays on Unemployment Duration and Programme Evaluation
150. **Sandén, Klas** (2006), Essays on the Skill Premium
151. **Deng, Daniel** (2006), Three Essays on Electricity Spot and Financial Derivative Prices at the Nordic Power Exchange
152. **Gebreeyesus, Mulu** (2006), Essays on Firm Turnover, Growth, and Investment Behavior in Ethiopian Manufacturing
153. **Islam, Nizamul Md.** (2006), Essays on Labor Supply and Poverty: A Microeconometric Application
154. **Kjaer, Mats** (2006), Pricing of Some Path-Dependent Options on Equities and Commodities
155. **Shimeles, Abebe** (2006), Essays on Poverty, Risk and Consumption Dynamics in Ethiopia
156. **Larsson, Jan** (2006), Four Essays on Technology, Productivity and Environment
157. **Congdon Fors, Heather** (2006), Essays in Institutional and Development Economics
158. **Akpalu, Wisdom** (2006), Essays on Economics of Natural Resource Management and Experiments
159. **Daruvala, Dinky** (2006), Experimental Studies on Risk, Inequality and Relative Standing
160. **García, Jorge** (2007), Essays on Asymmetric Information and Environmental

Regulation through Disclosure

161. **Bezabih, Mintewab** (2007), Essays on Land Lease Markets, Productivity, Biodiversity, and Environmental Variability
162. **Visser, Martine** (2007), Fairness, Reciprocity and Inequality: Experimental Evidence from South Africa
163. **Holm, Louise** (2007), A Non-Stationary Perspective on the European and Swedish Business Cycle
164. **Herbertsson, Alexander** (2007), Pricing Portfolio Credit Derivatives
165. **Johansson, Anders C.** (2007), Essays in Empirical Finance: Volatility, Interdependencies, and Risk in Emerging Markets
166. **Ibáñez Díaz, Marcela** (2007), Social Dilemmas: The Role of Incentives, Norms and Institutions
167. **Ekbom, Anders** (2007), Economic Analysis of Soil Capital, Land Use and Agricultural Production in Kenya
168. **Sjöberg, Pål** (2007), Essays on Performance and Growth in Swedish Banking
169. **Palma Aguirre, Grisha Alexis** (2008), Explaining Earnings and Income Inequality in Chile
170. **Akay, Alpaslan** (2008), Essays on Microeconometrics and Immigrant Assimilation
171. **Carlsson, Evert** (2008), After Work – Investing for Retirement
172. **Munshi, Farzana** (2008), Essays on Globalization and Occupational Wages
173. **Tsakas, Elias** (2008), Essays on Epistemology and Evolutionary Game Theory
174. **Erlandzon, Karl** (2008), Retirement Planning: Portfolio Choice for Long-Term Investors
175. **Lampi, Elina** (2008), Individual Preferences, Choices, and Risk Perceptions – Survey Based Evidence
176. **Mitrut, Andreea** (2008), Four Essays on Interhousehold Transfers and Institutions in Post-Communist Romania
177. **Hansson, Gustav** (2008), Essays on Social Distance, Institutions, and Economic Growth
178. **Zikhali, Precious** (2008), Land Reform, Trust and Natural Resource Management in Africa
179. **Tengstam, Sven** (2008), Essays on Smallholder Diversification, Industry Location, Debt Relief, and Disability and Utility
180. **Boman, Anders** (2009), Geographic Labour Mobility – Causes and Consequences
181. **Qin, Ping** (2009), Risk, Relative Standing and Property Rights: Rural Household Decision-Making in China
182. **Wei, Jiegen** (2009), Essays in Climate Change and Forest Management
183. **Belu, Constantin** (2009), Essays on Efficiency Measurement and Corporate Social Responsibility
184. **Ahlerup, Pelle** (2009), Essays on Conflict, Institutions, and Ethnic Diversity
185. **Quiroga, Miguel** (2009), Microeconomic Policy for Development: Essays on Trade and Environment, Poverty and Education
186. **Zerfu, Daniel** (2010), Essays on Institutions and Economic Outcomes
187. **Wollbrant, Conny** (2010), Self-Control and Altruism
188. **Villegas Palacio, Clara** (2010), Formal and Informal Regulations: Enforcement and Compliance

189. **Maican, Florin** (2010), Essays in Industry Dynamics on Imperfectly Competitive Markets
190. **Jakobsson, Niklas** (2010), Laws, Attitudes and Public Policy
191. **Manescu, Cristiana** (2010), Economic Implications of Corporate Social Responsibility and Responsible Investments
192. **He, Haoran** (2010), Environmental and Behavioral Economics – Applications to China
193. **Andersson, Fredrik W.** (2011), Essays on Social Comparison
194. **Isaksson, Ann-Sofie** (2011), Essays on Institutions, Inequality and Development
195. **Pham, Khanh Nam** (2011), Prosocial Behavior, Social Interaction and Development: Experimental Evidence from Vietnam
196. **Lindskog, Annika** (2011), Essays on Economic Behaviour: HIV/AIDS, Schooling, and Inequality