GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

The link between large-scale land acquisition and fertility: an empirical investigation in sub-Saharan Africa

Master thesis
Faculty of Business and Economics
University of Göttingen

First supervisor: Prof Dr Renate Hartwig
Second supervisor: Rebecca Lohmann

Author: Bianca Dülken
Student number: 21740701
Master: International Economics
Place of birth: Bergisch Gladbach

Date of submission: April 12, 2021

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

DiD Difference-in-Differences

DHS Demographic and Health Surveys

GIGA German Institute for Global and Area Studies

IPV Intimate Partner Violence

LARCC Large-scale Land Acquisitions, Rural Change and Social Conflict

LSLA Large-scale land acquisition

OLS Ordinary least squares

PCA Principal component analysis

SSA Sub-Saharan Africa

USAID United States Agency for International Development

1 Introduction

In sub-Saharan Africa (SSA), the number of births per woman[1] is steadily declining. In comparison to other regions of the world, however, birth rates are at a persistently high level (see Figure A2) and within SSA there are large differences in fertility rates (Bongaarts and Casterline, 2013; World Bank, 2021a). On a macro level, high fertility combined with declining mortality has led to rapid population growth – constantly above 2.5 percent – that poses social, economic, and environmental challenges and hinders sub-Saharan African countries in raising their standard of living (Bongaarts and Casterline, 2013; Kubitza and Gehrke, 2018; World Bank, 2021b). By the year 2100, SSA's population is expected to be tripled accounting for 3.07 billion people (Ezeh et al., 2020).

Due to rapid population growth, arable land is becoming scarce in large parts of the African continent. As land access is important for many people to secure their livelihoods, land scarcity has become a driver for internal migration, especially among young men (Holden and Otsuka, 2014). To find farmland, young men might migrate to areas surrounded by large-scale land acquisitions[2] (LSLAs) as several studies report in-migration to these areas (Deininger and Byerlee, 2012; Gyapong, 2020; Herrmann, 2017). A study on land rights in Liberia provides anecdotal evidence that men who migrate to an area for long-term settlement can access communal land through marrying a local woman (Dodd et al., 2018). Thus, it is imaginable that young men migrate to areas surrounded by LSLAs and marry a local woman to gain land access. The increasing likelihood of marriage is expected to result in a higher fertility rate (Kubitza and Gehrke, 2018). Consequently, through the migration-marriage linkage, the demand for children in areas surrounded by LSLAs is expected to rise.

However, several studies provide evidence that due to a lack of employment or farming opportunities young men rather leave areas surrounded by LSLAs (Kleemann and Thiele, 2015; Ryan, 2018; Schoneveld, 2017). Moreover, qualitative findings of a recently conducted telephone survey in ru-

[1]According to World Bank (2021a), the total fertility rate is calculated as the number of children a woman would have born if she lived to the end of her childbearing years and gave birth according to the age-specific fertility rates for the year indicated.

[2]Following the Land Matrix (2020a), a land deal is defined as 'any intended, concluded, or failed attempt to acquire land through purchase, lease, or concession for agricultural production, timber extraction, carbon trading, industry, renewable energy production, conservation, and tourism in low- and middle-income countries'. Following the terms 'large scale land deals', 'large scale land investments', 'large scale land acquisitions' and 'LSLAs' will be used interchangeably.

ral sites of Liberia show that migrants can obtain access to communal land predominantly through the permission of the village chief or elders, rather than through marrying a local woman (see Table A11).[3] Hence, the migration-marriage linkage might not be a key driver of fertility increases in areas surrounded by large-scale land deals.

Contrary to the previous argument, the telephone survey conducted in Liberia provides suggestive evidence for a negative impact of LSLA proximity on individual fertility. Out of 84 Liberian women who are living nearby one of the two largest oil palm investments (henceforth called 'plantations') in the country, 32 women report a decreasing child demand since the plantations have been established (see Table A11).[4] Moreover, at the global level, ten of the top twenty LSLA target countries identified in Nolte et al. (2016)'s report are located in SSA. Using data provided by the World Bank (2021a) allows for comparing the fertility rate of these ten countries with the average fertility rate of SSA countries. It can be observed that for eight of these ten countries, the average number of births per woman is below the SSA average (see Table A6). Ceteris paribus, this might provide first – even though very rudimental – evidence for a negative linkage between LSLAs agglomeration and fertility.

Women's health burden (including the risk of dying in childbirth) and their socio-economic well-being are directly linked to declining fertility (Campbell et al., 2006; Chen et al., 1974; Miller, 2010). Therefore, in demographic literature several theories about what triggers decreasing fertility rates exist. Examples are studies analysing household income effects, changes in maternal opportunity costs of time or the degree of female empowerment (Kubitza and Gehrke, 2018). The question arises which role these three mechanisms – household income, maternal opportunity costs of time and female empowerment – play in explaining the LSLAs-fertility linkage.

Thereby, the influence of LSLA proximity on the livelihoods of local women and men is an intensively and controversially debated topic in academic literature and thus ambiguous impacts on the three mechanisms are expected. The term 'land grab' has been published numerous times and describes negative impacts such as loss of access to land and resources that could engender socio-economic deterioration (Cotula et al., 2009). While some researches support these findings (Li, 2011;

[3]The survey was under supervision of the 'Large-scale Land Acquisitions, Rural Change and Social Conflict' (LARCC) research team of the German Institute for Global and Area Studies (GIGA). More information about the survey design is provided in subsection A.3 of the data appendix.

[4]As expected, most women have reported that child demand stayed unchanged (40 out of 84 women).

Nolte and Ostermeier, 2017), there are also studies underlining potential opportunities for the people living close to a LSLA especially regarding employment creation and rural development (Deininger and Byerlee, 2012; Herrmann, 2017; Van den Broeck et al., 2017). Analysing the unexplored linkage of LSLA proximity and individual fertility might therefore add new arguments to the longstanding debate over whether LSLAs benefit or rather harm people living close to LSLAs and in this case women in particular. Hence, the research question of this thesis is whether and through which mechanisms the geographical proximity to large-scale land investment sites affects individual fertility in SSA.

In this thesis, a repeated cross-sectional dataset including individual data of women living in SSA countries geographically matched with locations of land deals will be used. Accounting for different implementation statuses of LSLAs, a geographic difference-in-differences estimation is applied as the main identification strategy. Thereby, the effect will be disentangling subsequently by examining the impact of potential mechanisms - household income, maternal opportunity costs of time and female empowerment. Anecdotal evidence from the telephone survey in Liberia underpins the analysis.

The thesis is structured as follows: In the second section the conceptual framework will be presented describing the demand for children as well as the selected mechanisms affected by LSLA proximity and altering fertility desires and reproductive outcomes. Section 3 describes the data and variables chosen for the quantitative analysis. In section 4, the applied estimation strategy will be explained in detail followed by some descriptive statistics (section 5). Section 6 will present the estimation results including a heterogeneity analysis and robustness checks. In section 7 caveats of the study design and the empirical results will be identified, and Section 8 concludes.

2 Conceptual framework

The conceptual framework of this analysis is based on Becker and Lewis (1974)'s quantity-quality model on the demand of children (henceforth Q-Q model). First, the theoretical foundation of the Q-Q model will be explained. Second, the three mechanisms shaping fertility will be linked to LSLA proximity in order to identify how proximity to LSLAs may affect individual fertility.

2.1 Child demand

In Becker and Lewis (1974)'s Q-Q model slightly modified by Kubitza and Gehrke (2018), children are treated analogously to consumer durables (e.g. houses) and a household utility function of the form U(n, q, Z) is assumed. Thereby, n describes the quantity of children (i.e. number of children), q the child quality (i.e. proxied by the amount spent on each child at given prices) and Z the quantities of other commodities. The household utility is maximised subject to the following budget constraint:

$$I = p_n \times n + p_q \times q + p_c \times n \times q + \pi_z \times Z \tag{1}$$

In the equation, the notation I represents the total income of a household. The child quantity costs per unit (i.e. costs of having one additional child) are defined by p_n and multiplied by n (i.e. the number of children). These costs include opportunity cost of time of pregnancy, cost with respect to individual child rearing and all other monetary or psychic expenditures on children, which are largely independent of child quality. For instance, with increasing contraception costs, p_n will fall.

Moreover, irrespective of how many children a woman has (n), costs of child quality (p_q) are included in this equation and multiplied by child quality (q). An example for these child quality costs (p_q) are expenditures on reusable schoolbooks and clothes that are passed on from child to child. Becker and Lewis (1974) describe the component ($p_q \times q$) as having attributes of a 'family good' since all children can benefit from the provision of the good.

Additionally to p_q which does not depend on n, costs of one unit of child quality (p_c) for one child are added. Examples for these quality costs, which are depending on n, are tuition fees or expenditures on nutrition. Thus, costs of child quality are disaggregated by whether they depend on n (p_c) or not (p_q). It is important to notice that Becker and Lewis (1974)'s framework only allows parents to increase child quality (q) by an identical amount for all children in the household. Thus, if parents decide to invest in child quality, they will raise quality for all children by one unit which gets more expensive for a larger n. This non-linear part of the household budget constraint ($p_c \times n \times q$) is of particular interest as it shows that, while child quantity and quality are entering separately in utility, they are closely connected through the household's budget constraint. Lastly, π_z represents the price for other commodities which are not related to cost regarding child quantity or quality (Becker and Lewis, 1974; Kubitza and Gehrke, 2018).

Studying the demand for children, the above-mentioned maximisation of the utility function has been cited numerous times. Yet, the assumptions of Becker and Lewis (1974)' neoclassic model are highly criticised. Studies based on bargaining models reject the assumption that a household maximises a single welfare function and that household members share the same preferences or that a dominant decision maker within the household exists. Furthermore, the unitary assumption that aggregated household resources alter fertility is rejected. Instead, it is advocated that individual preferences, resources and income sources of the household members be considered in studies analysing the impact on fertility (Field, 2003).

2.2 Mechanisms – Child demand and land deals

Household income, maternal opportunity costs of time as well as female empowerment are considered as potential pathways through which women's fertility desires and reproductive outcomes are determined. Changes in fertility caused by alteration in household income or maternal opportunity costs of time can be explained by the Q-Q model. Instead, consistent with the Q-Q model critique described in the previous subsection, the impact of female empowerment on fertility is based on women's autonomy and intra-household bargaining power rather than unitary household decisions.

These three mechanisms are assumed to be altered by the establishment of LSLAs in SSA countries. Thus, the mechanisms are expected to be the link for the association between LSLA proximity and fertility.[5]

Household Income: An increase in household income could be invested in increasing the number of children, increasing child quality or both.[6] Consequently, it is possible that the demand for children increases or decreases when income changes (Kubitza and Gehrke, 2018). Following Becker and Lewis (1974)'s analysis, the income elasticity of child quality is assumed to be higher than the income

[5] Due to the limited scope of this thesis and data limitations, not all potential mechanisms will be analysed. Further research possibilities will be mentioned in section 7.

[6] Child quantity and quality are assumed to be normal goods. If the absolute income increases, the demand for child quality and quantity rises (Doepke, 2015).

elasticity of child quantity.[7] Through the interaction term between quality and quantity (see Equation (1)), an increase in the quality of each child affects the shadow price of child quantity. Hence, even a small increase in child quality (q) could have a negative and large impact on demand for children. Findings from the Liberia telephone survey support this theoretical argument as most women stated that they do not like to have another child since they rather prefer to invest in the education of their children (see Table A11). The income effect on child quantity is therefore expected to be negative (Kubitza and Gehrke, 2018).[8] Hence, the question arises whether the establishment of LSLAs leads to an increase or a decrease of household income of the people surrounded by LSLAs.

Findings from academic research show that access to employment on LSLAs increases household income substantially (Herrmann and Grote, 2015; Herrmann, 2017; Maertens et al., 2011). A study analysing the Senegalese tomato sector, for instance, provides evidence that household incomes of daily or seasonal workers were significantly higher than those for non-worker households (Maertens et al., 2011). Studies examining the impacts of LSLAs on income from sugarcane production in Malawi and Tanzania have been reported similar findings (Herrmann and Grote, 2015; Herrmann, 2017).

Nevertheless, as production of LSLAs is typically capital-intensive and thus capacities to absorb local employment are limited (Nolte et al., 2016), the net effect of LSLAs on employment is likely to be negative (Nolte and Ostermeier, 2017). Income gains from employment on LSLAs may not compensate for income losses in labour-intensive smallholder agriculture if smallholders or pastoralists lose land access due to the LSLA establishments (Baumert et al., 2019; Nolte and Ostermeier, 2017). Thereby, the common perception that LSLAs are predominantly targeted at idle land (Deininger and Byerlee, 2012), has been disproved (Cotula et al., 2009). Instead, easily accessible croplands, remote woodlands, and moderately accessible shrub- or grasslands typically used by smallholders and pastoralists are targeted. Consequently, competition over land in areas targeted by LSLAs increases

[7] The expected high income elasticity of child quality is based on an analogy with other durable consumer goods. For instance, as households become richer, they do not tend to buy large numbers of houses, but rather opt for higher quality, e.g. a house with more bedrooms (Doepke, 2015).

[8] The validity of the negative fertility-income relationship has been debated a lot. Opponents claim that the negative relationship is nothing more than a statistical fluke due to a missing variable problem (Jones et al., 2008). Since this debate is not within the scope of this thesis, it is assumed that an increase in income has a positive effect on child quantity and quality. Thereby, due to the substitution effect and the higher income elasticity of child quality, child quantity decreases if income increases.

whereby rural households are likely to lose access or even ownership of land (Messerli et al., 2014; Nolte and Ostermeier, 2017). While data and empirical evidence is scarce (Kleemann and Thiele, 2015), losing access to land eventually accompanied by inadequate compensation payments will have direct, negative income effects as documented by some qualitative studies (Kleemann and Thiele, 2015; Nolte et al., 2016; Schoneveld et al., 2011; Schoneveld, 2017).

Inclusive business models such as contract farming[9] can prevent crowding out of smallholder farmers (Nolte and Ostermeier, 2017) and are considered to have a greater, positive effect on household income than mere plantation employment (Herrmann and Grote, 2015; Herrmann, 2017). Yet, contract farming schemes are only offered by a minority of investors (Kleemann and Thiele, 2015).

Since the establishment of LSLAs is associated with greater capital inflows into surrounding areas, income levels of not directly affected households might be altered as well. Spillovers from LSLAs to the (smallholder) farming sector through technology transfer, improved infrastructure, and access to markets for inputs, output and credit may occur (Collier and Venables, 2011; Nolte and Ostermeier, 2017). Recent empirical studies have, however, shown that the evidence for such spillovers is rather limited (Ali et al., 2019; Deininger and Xia, 2016; Lay et al., 2018). Yet, besides benefiting the farming sector, the inflow of capital might change the sectoral composition from employment in agriculture to employment in industry and services (Dorward, 2013; Nolte and Ostermeier, 2017). Productivity gains are thus likely to lead to higher total labour incomes of people living nearby LSLAs (Food and Agriculture Organization (FAO), 2016).

To sum up, based on the Q-Q model a negative income-fertility relationship is assumed. Taking into account the above-mentioned channels – LSLAs employment, land dispossession, and spillovers – the direct income effect of the arrival of LSLAs is expected to be negative whereby indirect, positive income effects are conceivable.

Maternal opportunity costs of time: Through female labour force participation, women are expected to reallocate their time from child rearing to income earning activities. According to the Q-Q model, p_n rises and, ceteris paribus, the demand for children (n) decreases (Kubitza and Gehrke, 2018).

[9] Contract farming schemes are pre-arranged supply contracts between local farmers and investors for the supply of agricultural products (Land Matrix, 2020a). Typically, investors provide local farmers with inputs such as improved seeds, fertiliser, machinery, or technical know-how (Behrman et al., 2012).

Out of 30 village chiefs being interviewed in the telephone survey in Liberia, 19 women state that local women are working on the plantations. Furthermore, family work seems to be common in some areas as more than half of the chiefs reported that – in case their husband works on the plantation – women help their husbands with the work. Analysing the relationship between female labour participation and maternal opportunity costs of time, the anecdotal evidence shows that working women have less time for child-rearing: Out of 15 chiefs who reported that either female labour participation on the plantation or in other jobs (excluding domestic work) has increased recently, 12 chiefs have reported that women have less time for child-rearing since the plantation was established (see Table A11). Quantitative evidence of whether women find employment on large-scale land farms, however, is rather limited and as stated above, the net effect of LSLAs on employment is expected to be negative. Consequently, if female labour is not sufficiently demanded within the area, female labour participation as well as female wages decrease. The resulting decrease in maternal opportunity costs of child rearing (p_n) is expected to increase fertility (n) (Kubitza and Gehrke, 2018).

However, if LSLAs establishment triggers local economic development, female labour demand in other sectors might rise (Kubitza and Gehrke, 2018). Besides working on the plantation, daily housework and care-giving, 15 out of 26 village chiefs being interviewed in the Liberia survey have reported an increase in female labour participation in the last couple of years (see Table A11). As female labour participation increases, p_n is expected to rise, leading to a lower child demand (n).

Moreover, LSLAs promoting economic development might lead to the creation of more jobs in the high education sector. If women are entering a more profitable sector, p_n would rise and consequently child demand (n) would decrease. Moreover, as wages between the low and high education sector widen, returns to women's education increase. Hence, women are expected to reallocate their time away from child rearing to invest in their education. A case study in Indonesia confirms these arguments since – due to the oil palm cultivation – women opted out of agriculture into the service sector and invested more time in their education. As maternal opportunity costs of time and thus p_n have increased, the demand for children (n) decreased (Kubitza and Gehrke, 2018).

To put it in a nutshell, female labour participation and women's education are negatively associated with fertility (Bongaarts, 2010). Since evidence is scarce, the influence of LSLAs on maternal opportunity costs of time remains ambiguous and thus an increase or decrease on fertility is imaginable.

Female empowerment: In his book, Sen (1999) argues that women's empowerment[10] is a key driver to reduce fertility rates. Regarding LSLA proximity, changes in women's land rights and usage, their participation in negotiations about land distribution and respective compensation payments with LSLAs investors as well as women's current occupation status are, among others, all factors potentially influencing women's female empowerment and thus their reproductive outcomes (Elmhirst et al., 2017; Kubitza and Gehrke, 2018).

A case study conducted in Indonesia shows that prior to large-scale oil palm cultivation, women's status in the family and community stemmed from their active role in swidden rice cultivation. Even if women did not formally own the land, their labour was valued, and women derived symbolic standing. During negotiations with the oil palm companies about land including swiddens, managed and used by women formerly, women were excluded from any decision-making processes (Elmhirst et al., 2017). Moreover, findings from case studies analysing women's land rights in Liberia show that compensation payments to women were either pending, not sufficiently high, or the compensation was given to the man and never reached the woman (Dodd et al., 2018). These factors weaken women's economic independence and might affect woman's household decision-making power negatively. Assuming that women have *per se* lower fertility preferences than men, the decreased intra-household bargaining weight of women would lead to an increase of child demand (Kubitza and Gehrke, 2018).

Yet, a reverse relationship of LSLAs establishments strengthening female empowerment is also imaginable. Considering female employment effects as described above, women might work on the LSLAs itself or find lucrative jobs in other sectors. A decreasing effect on fertility would then simply reflect the increased intra-household bargaining power of women as women start earning their own income over which they have higher control than farm income (Heath and Mushfiq Mobarak, 2015; Kubitza and Gehrke, 2018).

[10]Following Sen and Batliwala (2000), female empowerment is defined as a process by which women gain greater control over circumstances of their lives. Besides compromising control over resources, the definition includes control over ideology. Thus, female empowerment leads to a greater self-confidence, and an inner transformation of consciousness that enables women to overcome external barrier. While Sen and Batliwala (2000) defined empowerment broadly, here specifically female empowerment is discussed.

In summary, a higher degree of female empowerment is expected to have a negative impact on child demand. In line with the inconclusive findings for the other two mechanisms, the impact of LSLA proximity on female empowerment and thus its linkage to fertility remains ambiguous.

3 Data

In this section, the data and variables used for the following quantitative analysis will be presented. The empirical analysis is based on a geographical match of spatial data of LSLAs accessed through the Land Matrix data platform with individual record data provided by the Demographic and Health Surveys (DHS).

3.1 Land Matrix

For large-scale land deals, data from the most comprehensive database on land deals, the Land Matrix, is used (retrieved on August 2020). The Land Matrix Global Observatory is an independent monitoring initiative launched in 2012 which collects, provides, and analyses data on land acquisitions. The database includes large-scale land deals covering 200 hectares and more that have occurred since the year 2000 (Land Matrix, 2020a). On a global scale, the Land Matrix has information about 2,018 deals covering an area of approximately 68 million hectares. Out of these, nearly one third (635 deals) are located in Africa. With concluded deals covering currently more than 13.5 million hectares of Sub-Saharan African land leased to foreign firms, this world region is globally the most targeted in the rush for land and therefore of particular interest when analysing socio-economic impacts of land deals (Land Matrix, 2020b).[11]

The Land Matrix categorises land deals into four implementation statuses: (1) not yet started, (2) start-up phase, (3) in operation, and (4) abandoned (Land Matrix, 2020a). For the empirical analysis, it is essential to differentiate between 'active' LSLAs being currently in operation (category (3)), 'inactive' LSLAs intended to be implemented in the future but not yet in operation stages (categories (1) and (2)) and LSLAs being 'abandoned' (category (4)).

[11] The figures were retrieved from the open-access, online Land Matrix database in August 2020.

In total, the initial sample of this thesis includes 289 land deals with any information on location and implementation status and excluding deals with the most inaccurate degree of spatial accuracy.[12] Since some land deals include multiple sub-locations, a total of 367 locations of land deals will be considered.

3.2 DHS

To analyse fertility desires and reproductive outcomes, micro data from the DHS Programme funded by the United States Agency for International Development (USAID) is being used. The standardisation of surveys across years and countries allows to build a repeated cross-sectional dataset. The dataset combines the women's questionnaire from 70 surveys conducted in SSA containing, inter alia, information about fertility, household bargaining power, further socio-economic information as well as GPS coordinates. To ensure respondent's confidentiality, DHS displaces the exact location of households in all surveys randomly. Urban clusters contain a positional error of maximal two kilometres (km). In rural areas the error is set up to five km with one percent of rural clusters being displaced a maximum of ten km. Thereby, the displacement is restricted to locations within the country and DHS survey region (ICF, 2021). The initial dataset contains 903,259 women of reproductive age (15–49 years old) from 32 SSA countries. The women were interviewed during 2000–2019[13] and live in 34,681 survey clusters[14] in 369 sub-national regions.

Studies on desired or realised fertility – such as Kiser and Hossain (2018)'s study on socio-economic impacts on fertility – often restrict their sample by exclusively including currently married women. However, these studies may not capture the entire effect of socio-economic variables on fertility since they do not take into account delayed marriage or contraceptive use (before marriage). Additionally, assuming marriage and child rearing to be a joint decision, studies based on married women may suffer from a potential sample selection bias favouring women with a higher demand for children (Ainsworth et al., 1996). To avoid these drawbacks, the baseline sample includes all women

[12] More details on the data provided by the Land Matrix is given in subsection A.3 of the data appendix.

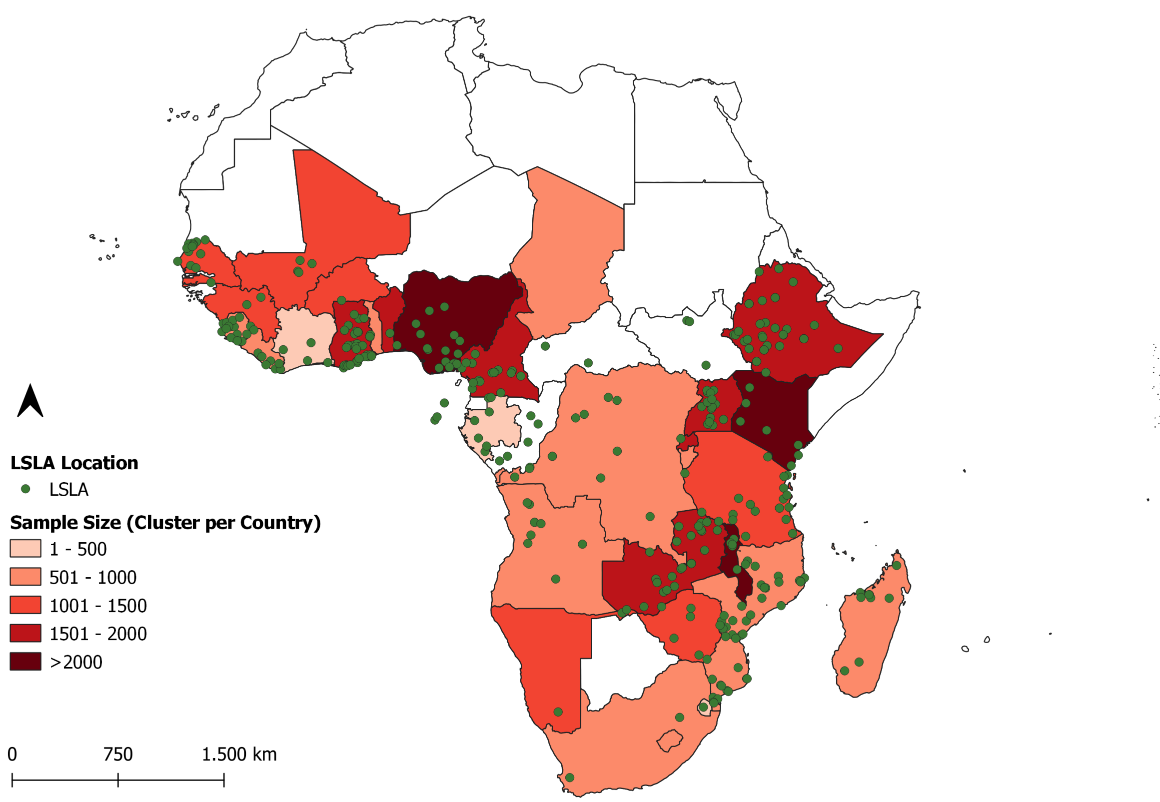
[13] For the year 2002, no survey data is available.

[14] Most DHS surveys cover about 25-30 households per cluster (Croft et al., 2018).

while differences in their marital status will be controlled for. Furthermore, a heterogeneity analysis will be applied by exclusively analysing LSLAs impact on fertility using DHS couples surveys.

In Figure 1, the distribution of LSLAs and DHS clusters included in this research is shown. Countries with geocoded DHS data are coloured red. The intensity of the colour indicates the number of clusters per country. As can be seen, the number of clusters per country is especially high in Kenya, Malawi, and Nigeria. The locations of large-scale land deals across SSA are presented as point coordinates. The map shows that in some regions – for instance in the southeast of Nigeria – agglomerations of LSLAs can be observed. In line with Lay and Nolte (2018) findings, this might be a hint for a non-random distribution of LSLAs.

Figure 1: LSLA locations and number of DHS clusters per country



(Own visualisation inspired by Kotsadam and Tolonen (2016))

Desired and realised fertility

In this thesis, three fertility outcomes are considered namely desired fertility, i.e. ideal number of children, and realised fertility, i.e. number of children ever born and currently fertility. Regarding realised fertility, analysing only the total number of children born per woman, a fundamental bias of women having already completed their fertility before the nearby LSLA was established could drive the findings of this thesis (Kubitza and Gehrke, 2018). To avoid this bias, a dichotomous variable measuring current fertility with a specific reference period of 12 months will be considered. This variable accounts for whether a woman has given birth and/or being pregnant within the last 12 months before the interview.

Mechanisms

Household Income: For measuring the mediating effect of household's economic status, wealth is chosen as a proxy for household income as it can be constructed by using DHS data. Moreover, wealth appears to be a suitable proxy for economic status in developing countries because, compared to alternatives such as household income and consumption expenditure, wealth information is less volatile. Thus, wealth represents a more permanent status (Rutstein and Johnson, 2004).

While the standard DHS wealth index is missing for some surveys and is elementary for this thesis, a wealth score is constructed following the guidance of Rutstein and Johnson (2004). Using a simple principal component analysis (PCA), the wealth score is calculated for each household considering the availability and quality of several household assets. Due to limited data availability, the score only captures selected assets and thus might not present differences in wealth to its full extend.[15]

Maternal opportunity costs of time: In the empirical part of this thesis, a binary variable indicating whether the respondent (i.e. the woman) is currently working and a binary variable measuring whether the respondent is working in a non-agriculture job will be considered. Yet, work performed by women in developing countries is often unpaid and informal (Bongaarts et al., 2019). This might affect maternal opportunity costs of time and is not covered by these variables. Moreover, no information about whether women work on the large-scale land investment sites is given.

[15] A more detailed description of the wealth score computation and its shortages is given in the subsection A.3 of the data appendix.

By including a woman's highest level of education attained – (1) no education, (2) primary education, (3) secondary education and (4) higher education – , the impact of women's education on maternal opportunity costs of time as described in the previous section will be considered.

Female empowerment: A widely accepted proxy for female empowerment using DHS data is women's participation in household decisions-making (Annan et al., 2021; Kishor and Subaiya, 2008; Upadhyay et al., 2014). A recent study by Annan et al. (2021) captures a novel dimension of women's empowerment considering intra-household contention over decision-making when analysing women's decision power relative to the power of their husbands. However, since DHS data on women's participation in decision-making is exclusively provided for currently married women, this proxy cannot be applied in the main analysis of this thesis. Alternatively, DHS information on attitudes towards physical intimate partner violence (IPV) against women is recognised as a proxy for female empowerment and will be applied. An index was calculated measuring whether the woman answered that intimate partner violence by the husband was not justified in any of the five hypothetical scenarios.[16] The index is positively associated with female empowerment and conscientisation in favour of gender equality (Kishor and Subaiya, 2008). However, the interpretation of this index must be done with a certain degree of caution. For instance, Schuler et al. (2011)'s study provides qualitative evidence that DHS standard questions do not accurately represent the proportion of women who condone IPV because of a lack of context and consideration of cultural and gender norms. Moreover, the multidimensional nature of female empowerment might not be captured by this proxy (Kishor and Subaiya, 2008). Nevertheless, based on data availability, the index measuring attitudes towards IPV against women will henceforth be used as (an imperfect) proxy for female empowerment.

Socio-demographic characteristics

To increase the goodness-of-fit of the following analysis, several socio-demographic characteristics likely to explain variance in fertility desires and reproductive outcomes are considered. Based on existing literature, respondent's current age, type of residence, marital status, the number of household members, the sex of the household head as well as religious affiliation will be included in the analysis

[16] The scenarios are (a) wife goes out without telling her husband, (b) wife neglects the children, (c) wife argues with her husband, (d) wife refuses to have sex with her husband, and (e) wife burns the food.

(Haq et al., 2017; Upadhyay and Karasek, 2012). Moreover, husband's highest level of education attained as well as the age of first cohabitation will be included as additional control variables in the heterogeneity analysis when the sample is restricted to couples surveys only.

When measuring the impact on respondent's ideal number of children, the number of living children is controlled for additionally (Upadhyay and Karasek, 2012). Furthermore, women's ideal number of children is expected to be affected by her husband's ideal number of children (Upadhyay and Karasek, 2012). Thus, husband's ideal number of children will be included in the heterogeneity analysis when the sample is restricted to couples surveys only.

For the measure of reproductive outcomes – number of children ever born and current fertility – the usage of modern contraceptives is controlled for in addition. As is widely recognised, the use of family planning products is key to preventing unintended pregnancies (Bongaarts and Casterline, 2018; Easterlin, 1975). Furthermore, the number of children who have died are included as another control variable when measuring the impact on the number of children ever born. It is assumed that if more children survive, fertility decreases (Kirk, 1996; Kubitza and Gehrke, 2018). In the telephone survey in Liberia, only one woman stated that she wants to have another child because her previous child died (see Table A11). Consequently, the loss of a child might not be the main driver for wanting another child.

4 Estimation strategy

4.1 Difference-in-differences

Combining several waves of DHS survey data with information on LSLAs, comparisons over both time and geographical location are possible. Outcomes of women living close to operating LSLAs and women living close to LSLAs operating in the future will be compared with women living further away from LSLAs. Hence, this thesis relies on a spatial– temporal estimation strategy and, following Kotsadam and Tolonen (2016)[17], a difference-in-differences method will be used.

[17] Also applied in other papers such as Knutsen et al. (2017) and Kotsadam et al. (2018).

As described in subsection 3.1, the classification of a land deal ('active', 'inactive' or 'abandoned') is based on its implementation status. By converting the Land Matrix dataset into panel data, it is possible to define the current implementation status of land deals for each year from 2000 to 2020. Thus, the implementation status of each matched LSLA at the time of the respective DHS interview is defined.

Following the approach of Kotsadam and Tolonen (2016) and assuming that fertility is affected within a cut-off distance of 25 km, three groups are defined: (1) a post-treatment group, which contains women living within 25 km from at least one 'active' LSLA[18], (2) a pre-treatment group, which are those women living within 25 km from at least one 'inactive' (but no 'active') LSLA[19], and (3) a 'control' group which includes women living more than 25 km away from any LSLA. Women being matched with only 'abandoned' LSLAs within the 25 km cut-off distance cannot be assigned to one of the three groups - pre-treatment, post-treatment and control. They are excluded from the main regressions.[20] Regardless of the distance, every woman is merged with at least one LSLA location while multiple matching is possible. The baseline regression equation is compiled as follows:

$$Y_{ivt} = \beta_1 \times active + \beta_2 \times inactive + \alpha_r + g_t + \lambda X_i + \varepsilon_{ivt} \tag{2}$$

Thereby, the fertility desire or reproductive outcome Y for an individual i, cluster v and year t is regressed on 'active' and 'inactive' binary, explanatory variables. All regressions control for region (α_r) and year (g_t) fixed effects. The vector X represents several independent and control variables grouped by: (1) socio-demographics, (2) wealth, (3) maternal opportunity costs of time, and (4) female empowerment. Variables are assigned to each group as explained in subsection 3.2. Considering that observations are not independent within each DHS cluster, robust standard errors are clustered at the level of the primary sampling unit (Kotsadam et al., 2018).

For all regressions, a difference-in-differences (DiD) estimator is provided to cover the difference ($\beta_1 - \beta_2$) between the 'active' and 'inactive' independent variable (Kotsadam and Tolonen, 2016).

[18] The matching strategy relies on year information. Thus, the 'active' variable includes proximity to land deals who have started operating at least the year before the DHS interview. If a land deal has started within the interview year, the matched woman is coded as living in an 'inactive' area.

[19] 'Inactive' LSLAs have not started operating before the time of the survey but are expected to operate in the future.

[20] The sample size is reduced by 4,839 women.

Thus, the equation does not solely consider differences between areas nearby LSLAs and areas far-away from any LSLA. The first variable, *active*, captures the difference in fertility between women living close to at least one operating LSLA and those living further away. The second variable, *inactive*, captures the difference in fertility between women living close to a future, not yet operating LSLAs and women living further away. A 'control' woman will therefore often enter the comparison group for both – pre- and post-treatment women (Knutsen et al., 2017). In line with the standard DiD setup, four groups (controlled and treated each in the pre- and post-period) are included (Wooldridge, 2020).

Interpreting the coefficient for 'active' in isolation would neglect that other variables might have affected the outcomes. Yet, LSLA's investment decisions could be influenced by pre-existing, socio-demographic structures. Examples for these pre-existing structures are population density, wage labour, infrastructure, or accessibility of inputs in that geographical location (Kotsadam and Tolonen, 2016). Thus, differences in socio-demographic characteristics of treatment and control areas which are not related to the LSLA operation directly are imaginable. As shown in Figure 1 and argued by Lay and Nolte (2018), LSLAs are expected to be distributed non-randomly. Including the 'inactive' dummy variable in the equation above, a potential selection bias by LSLA location choice is controlled for. If such a bias exists, the DiD estimator needs to be considered. The DiD approach thus represents the main identification strategy of this thesis.

4.2 Limited dependent variables

For all three dependent variables, a simple ordinary least squares (OLS) regression might not provide unbiased results. First, the variable measuring respondent's ideal number of children ('fertility preference') suffers from right-censoring. Only responses up to six children are given. For responses greater than that, it is known that the answer is at least equal to six children. Using this censored variable, the highest proportion of women reported an ideal number of children of six or more (see Figure A5). Following Wooldridge (2020), a Tobit regression was constructed taking censoring into account.

Second, about 27 percent of the women in the sample have had no live births (see Figure A4). Taking censoring at zero into account, least square regression coefficients are inconsistent and – under the normal distribution assumption – biased downward proportional to the degree of censoring. The Tobit model for a corner solution accounts for this bias and will thus be used for measuring LSLAs impact on the number of children ever born (Ainsworth et al., 1996; Wooldridge, 2020).

Last, the maximum likelihood estimation of Logit or Probit models is appropriate for nonlinear binary responses. Therefore, the Probit model is chosen for the model measuring the impact on current fertility (Wooldridge, 2020).

4.3 Distance cut-off

Based on previous spatial analyses using similar data (see Knutsen et al., 2017; Kotsadam and Tolonen, 2016; Kotsadam et al., 2018), a baseline distance of 25 km from the LSLAs is selected. This distance cut-off is chosen considering the following reasons: (1) The DHS locations are randomly displaced as explained in subsection 3.2 and thus small cut-off distances are likely to introduce more noise. Moreover, (2) LSLAs geocoordinates reflect the centroid of the investment area – information on boundaries capturing the whole investment area is not given. When choosing a rather narrow distance cut-off, the likelihood of capturing the actual LSLAs area instead of the surrounding villages increases. (3) Increasing the distance cut-off leads to a rapidly increase of the sample size (see Table A9). All else equal, this would increase the validity of the results. However, using large distances harbours the risk of not capturing the LSLAs footprint anymore (Knutsen et al., 2017; Kotsadam and Tolonen, 2016). To balance potential attenuation biases arising from too short or too long distances, a baseline cut-off distance of 25 km from the LSLAs is chosen. It is acknowledged that the distance is still arbitrary to some extent. Thus, the robustness of the results will be tested by applying distance cut-offs of 15 and 50 km.

5 Descriptive analysis

Before moving on to the regression analysis, descriptive statistics will be presented. First, general characteristics of the sampled women and the balance of these characteristics among the three sub-samples – active, inactive and control – will be discussed using simple t-tests. Following, the mean differences of the dependent variables and independent variables of main interest between the three sub-samples are examined in detail. Statistics on the whole sample including the sample size as well as minimum and maximum values per variable are provided in Table A9 in the appendix.

5.1 Sample characteristics

Out of the initial sample with 367 LSLAs locations, 363 locations located in 31 countries are included in the final sample (see Table A7). All DHS clusters are matched to LSLAs, but four LSLAs are not matched to clusters. Three of the four not-matched LSLAs locations – green point coordinates in Figure A3 – are located on the islands of São Tomé und Príncipe and one land deal is situated in the Northern part of South-Sudan. As can be seen in Figure A3, the LSLAs deals are not nearby any DHS cluster – red point coordinates – and are dropped in the final sample.[21]. Following Kotsadam and Tolonen (2016), the largest cut-off distance is set to 200 km in this thesis and will be addressed in the robustness subsection 6.3. Moreover, out of the initial DHS sample accounting for 903,259 women, 8,183 women were merged with LSLAs being 'abandoned' after the year of the DHS interview. Since further information on the implementation status before the project was 'abandoned' is missing for these land deals, the 8,183 women could not be assigned to one of the three groups – 'active', 'inactive' and 'control'– and have been excluded from the analysis. Table A8 provides the number of DHS interviews disaggregated by country and years of the final sample accounting for 895,076 women.

In the final sample, respondents live on average 155 km away from the closest LSLAs. Considering the chosen distance cut-off of 25 km and the three groups defined, 2.58 percent of all women live within 25 km of 'active' LSLAs, 5.95 percent within 25 km of 'inactive' LSLAs and 0.54 percent within 25 km of 'abandoned' LSLAs. Thereby, the proportions increase significantly when the distance cut-offs are extended to cover 50 or 200 km (see Table A9).

In Table A10, the mean values of control variables are presented disaggregated by the three groups – control, active and inactive – using the 25 km cut-off distance (see columns (1) to (3)). Women being matched with 'abandoned' LSLAs within the 25 km distance are excluded. The p-values of

[21] For these four land deals, the closest DHS clusters are matched with other, closer land deals. In Figure A3, the second and third subfigures are enlarged regions of the first subfigure to provide evidence for the long distances between these four land deals and the closest DHS clusters.

mean comparisons between groups are reported in columns (4) to (6). Out of 813,879 women living in 'control' areas, 66 percent are located in rural regions. In comparison, women in 'active' areas are – on average – living in a more urban setting as only 51 percent are located in rural areas. Yet, the proportion is even smaller for 'inactive' areas accounting for 44 percent. The differences in proportions are statistically significant among the three groups.

Moreover, comparing the proportion of women who have never moved to another place in their life among the three sub-samples shows that in-migration to LSLAs areas might play a role. More specifically, 35 percent of the women in 'active' areas are living in this location their whole life whereas for 'control' areas the proportion is higher accounting for 48 percent. Again, differences are highly significant.

In all three sub-samples, women are - on average - between 28 - 29 years old. Furthermore, most women are currently married with the proportion being highest in the 'control' sample accounting for 56 percent and lowest in the 'active' group with 47 percent. The highest proportion of women never being in union is given in the 'inactive' group with 32 percent, followed by 31 percent in the 'active' group and 27 percent in the 'control' group. Following the positive marriage-fertility linkage stated in the introduction of this thesis, a higher proportion of women never been in union combined with a lower proportion of married women in the 'active' and the 'inactive' groups compared to the 'control' group might result in a lower fertility rate. Consistent with these findings, the proportion of women who had their first cohabitation before they were 15 years old is highest in the 'control' group and lowest in the 'active' group. In all three groups, about half of the respondents reported their age at first cohabitation within the age range of 15-19 years.

Regarding religious affiliation, for all three sub-samples most women are Christians, followed by Muslims. The difference in proportions for these categories is not statistically significant. The mean number of household members is significantly higher for households situated in 'active' and 'inactive' compared to 'control' areas. Yet, the magnitude is small and in all three groups the average accounts for approximately seven household members. Considering the sex of the household head, a significantly higher proportion of households in 'active' and 'inactive' areas is female headed (32 and 33 percent) compared to 'control' areas with female headed households accounting for 27 percent.

The potential out-migration of young men described in the introduction part of this thesis might explain why – on average – more households are female headed in areas surrounded by large-scale land investments.

Considering husband's characteristics, husband's ideal number of children is – on average – significantly lower in 'active' and 'inactive' areas (4.6 and 4.7) compared to the mean in 'control' areas (4.9). Yet, the difference is rather small. Comparisons of husbands' highest education level attended show that in the 'control' group most husbands did not attend any education (36 percent), while the highest proportion of husbands in the 'active' group attended secondary education (36 percent).

Regarding family planning, the proportion of women using modern contraceptives differs significantly between the sub-samples with the highest proportion in the 'active' group (23 percent) and lowest in the 'control' group with 19 percent. Yet, the differences are of rather small magnitude.

Lastly, the number of living children is highest among the 'control' group and lowest for the 'inactive' group accounting for a mean of 2.2 children. Furthermore, in 'active' areas the number of children died is significantly lower accounting for a mean of 0.3 children who have died per woman whereas the mean in the 'control' group accounts for 0.4.

The balance checks in Table A10 show that average values of the control variables differ between the three groups and that differences are often statistically significant. The significant differences between 'inactive' and 'control' areas could be an indication of a selection bias due to the location of the LSLAs, thus underlining the importance of the DiD estimation applied in this thesis. Nevertheless, since the highly significant p-values are likely to be driven by the large sample size (Wooldridge, 2020), the economic relevance of the statistically significant differences remains debatable. Moreover, the endogeneity problem of whether LSLAs investors target specific areas (e.g. rather urban locations) or due to the LSLAs establishment rural development is promoted cannot be resolved.

5.2 Fertility and land deals

Table 1 depicts the means of the fertility variables disaggregated by the three sub-samples. Thereby, in column (4) to (6) the p-values of the mean differences are given.

The number of children ever born per woman is with a mean of 2.6 significantly lower for women of the 'active' group compared to women of the 'control' group accounting for 2.9 children. Interestingly, the mean number of children ever born in all sub-samples is significantly lower than the mean ideal number of children a woman would like to have (4.2 - 4.4 children).[22] Considering the number of children a woman would have if she lived to the end of her childbearing years (see the average for SSA countries in Figure A2), the average number of children ever born is expected to be significantly higher. Yet, the sample is dominated mainly by women in their late twenties, who are likely to have not yet completed their fertility.

Table 1: Descriptive statistics for fertility variables

Variable	(1) Mean Control	(2) Mean Active	(3) Mean Inactive	(4) Active vs. Control	(5) Inactive vs. Control	(6) Active vs. Inactive
Ideal No of Children	4.361	4.184	4.167	0.000***	0.000***	0.574
No of Children Ever Born	2.876	2.576	2.536	0.000***	0.000***	0.284
Current Fertility	0.267	0.226	0.228	0.000***	0.000***	0.686
Observations	813,879	23,130	53,228	837,009	867,107	76,358

Notes: Columns (1) to (3) report the mean values disaggregated by the three groups – active, inactive and control – within 25km radius. Columns (4) to (6) report the p-values of the differences between means of the three groups. Women matched with abandoned LSLAs within 25 km are excluded. Robust standard errors are clustered at the level of DHS primary sampling unit. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

For all fertility variables, simple t-tests comparing the means confirm that – on average – fertility is significantly lower in targeted areas ('active' or 'inactive) compared to 'control' areas. For the two variables ideal number of children and the number of children ever born, the mean values are lower in 'inactive' compared to 'active' areas. Solely, for the variable considering a specific time span, i.e. current fertility, the mean of women living in 'active' areas is lower compared to the 'inactive' sub-sample. Yet, the differences between 'active' and 'inactive' areas are of small magnitude and statistically insignificant.

5.3 Fertility and mechanisms

Table 2 depicts the mean values and differences in means of the proxies for the three mechanisms discussed in the conceptual framework disaggregated by the sub-samples.

[22]Taking the censoring problem described in the previous section into account, the average ideal number of children is expected to be even higher.

Women living in ’active’ or ’inactive’ areas have – on average – a significantly higher wealth score than women of the ’control’ group. This might provide evidence for the existence of indirect income effects caused by LSLA proximity as explained in the conceptual framework. In line with this, the telephone survey conducted in Liberia reports a positive impact of the plantation establishments on rural development. 25 of all 30 village chiefs being interviewed mentioned changes such as the establishment of markets, schools, clinics or improvement of buildings and roads since the plantations have been established (see Table A11). However, the – on average – higher wealth score might also be linked to a LSLA location selection bias since the difference between ’inactive’ and ’control’ areas is significant. Moreover, it is noteworthy that the average score in all three groups is comparatively low, taken into account that the maximum wealth score accounts for ten (see Table A9).

Table 2: Descriptive statistics for mechanisms

Variable	(1) Mean Control	(2) Mean Active	(3) Mean Inactive	(4) Active vs. Control	(5) Inactive vs. Control	(6) Active vs. Inactive
Wealth Score	2.667	2.892	2.888	0.000***	0.000***	0.895
Currently Working	0.595	0.570	0.548	0.001***	0.000***	0.016**
Work Non-Agricultural	0.337	0.387	0.387	0.000***	0.000***	0.998
Education						
No Education	0.336	0.213	0.259	0.000***	0.000***	0.000***
Primary	0.340	0.366	0.312	0.001***	0.000***	0.000***
Secondary	0.282	0.368	0.375	0.000***	0.000***	0.551
Higher	0.042	0.053	0.055	0.002***	0.000***	0.513
Not condones IPV	0.514	0.568	0.517	0.000***	0.614	0.000***
Observations	813,879	23,130	53,228	837,009	867,107	76,358

Notes: Columns (1) to (3) report the mean values disaggregated by the three groups – active, inactive and control – within 25km radius. Columns (4) to (6) report the p-values of the differences between means of the three groups. Women matched with abandoned LSLAs within 25 km are excluded. Robust standard errors are clustered at the level of DHS primary sampling unit. **** p<0.01, ** p<0.05, * p<0.1.*

Considering the highest level of education attained, the proportion of women having no education is significantly higher in the ’control’ group than in the treatment groups (’active’ and ’inactive’). Accounting for 21 percent, the proportion is lowest for women living in ’active’ areas and the difference in proportions to women living in ’inactive’ areas is statistically significant. Compared to the other two groups, the share of women attended the primary education level is highest for women living in the ’active’ group. For the other two categories - secondary and higher - the difference in proportions between women in ’active’ and ’inactive’ areas (column (6)) is not significant and thus differences might not be linked to the operation of nearby LSLAs. However, for both sub-samples 'active' and 'inactive', the proportion of women having attended the secondary or higher education level is higher than in the 'control' group and its difference is statistically significant (columns (4) and (5)). This might provide evidence for higher maternal opportunity costs of time for women living in 'active' and 'inactive' areas as they - on average - have attained higher levels of education. However, the proportion of women currently working in 'active' areas (57 percent) is slightly lower compared to 'control' areas (60 percent) which might have the opposite effect on maternal opportunity costs of time. Considering female employment, the differences in proportions are all statistically significant and the proportion of women currently working is lowest for women living in 'inactive' areas (see column (3)). Since the difference between 'inactive' and 'control' areas is significant, the lower proportion of women currently working in treatment areas might be linked to employment effects occurring already in the pre-operating phase of the LSLAs. In line with that, the differences for 'active' and 'inactive' areas regarding women working in non-agricultural jobs is not significant. However, a significant difference of women working in non-agricultural jobs can be observed with a higher proportion in 'active' and 'inactive' areas (39 percent) compared to 'control' areas (34 percent). This might fit Kubitza and Gehrke (2018)'s argument that due to the LSLAs establishment some women opt out of agriculture into other sectors.

Considering the empowerment proxy, 57 percent of women living in 'active' areas state that IPV is not justified in any of the five scenarios, while in 'inactive' and 'control' areas the proportion is significantly lower (51 - 52 percent). Notably, the difference of proportions for 'inactive' and 'control' areas is not significant and highly significant between 'active' and 'inactive' regions. Thus, the difference in proportions of this proxy for female empowerment might be directly linked to the LSLA operation.

To sum up, proximity to LSLA is expected to have a negative impact on fertility. Moreover, the descriptive analysis shows that women in treatment areas (either 'active' or 'inactive') are on average wealthier[23], have a higher proportion of women working in the non-agricultural sector and have attained a higher level of education, as well as women reporting higher levels of female empowerment than women who do not live nearby any LSLA. For some proxies, differences between 'active' and

[23]Here, household wealth is considered.

’inactive’ areas turn out to be insignificant, while differences between ’inactive’ and ’control’ areas were mostly significant. This underlines the relevance of the DiD approach used in the upcoming section.

6 Estimation results

In this section, the results of the regression analysis – based on the estimation strategy explained in section 4 – will be discussed. For each dependent variable, findings from the simple OLS model will be examined before regression specifications are elaborated upon. Thus, the limitations of the dependent variables are taken into account. Through heterogeneity and robustness checks the validity of the main findings will be tested.

6.1 Main findings

Impact on ideal number of children

Table 3 presents the results from the simple OLS and Tobit model measuring the effect of LSLA proximity on fertility preferences, i.e. the ideal number of children stated by the respondents. Having a limited dependent variable, the interpretation of Tobit regression results may not be as straight forward as in the OLS model (Wooldridge, 2020). Taking the upper-censoring of the dependent variable (described in section 4) into account, the interest relies on interpreting the coefficients of the latent regression model. Thus, the coefficients of the dependent variable can be interpreted just as in a linear regression model (Wooldridge, 2020).

Applying the estimation equation described in section 4, the columns (1) to (7) of Table 3 show the results of the DiD model. The DiD estimator measures the impact on the ideal number of children of women living in ’active’ areas with at least one operating LSLAs within 25 km compared to ’inactive’ areas where at least one LSLA will operate in the future. In the case that LSLAs investors have chosen the location indiscriminately, there should be no statistically significant difference in fertility preferences in ’control’ and ’inactive’ areas, since in ’inactive’ areas the LSLA operation has not yet been started (Kotsadam et al., 2018).

Adding the proxy variables for the three mechanisms – household income, maternal opportunity costs of time and female empowerment – in the Tobit model one at a time (columns (4) to (6)), the interpretation of the inclusion of these variables follows standard mediation logic. A reduced coefficient on X (i.e. 'active' and 'inactive' coefficients) if M (i.e. proxies for one mechanism) is introduced into the regression indicates that part of the effect of X on the dependent variable Y is mediated through M (Knutsen et al., 2017).

Table 3: Effect of land deals on ideal number of children

	(1) OLS	(2) Tobit	(3) Tobit	(4) Tobit	(5) Tobit	(6) Tobit	(7) Tobit
Active 25 km	-0.134***	-0.259***	-0.156***	-0.136***	-0.141***	-0.148***	-0.106***
	(-5.61)	(-7.82)	(-6.40)	(-5.61)	(-5.60)	(-6.09)	(-4.26)
Inactive 25 km	-0.199***	-0.338***	-0.111***	-0.046**	-0.119***	-0.130***	-0.051**
	(-11.28)	(-13.86)	(-5.57)	(-2.13)	(-6.04)	(-6.68)	(-2.44)
Socio-Demographics	No	No	Yes	Yes	Yes	Yes	Yes
Household Income	No	No	No	Yes	No	No	Yes
Opportunity Costs	No	No	No	No	Yes	No	Yes
Female Empowerment	No	No	No	No	No	Yes	Yes
Difference in differences	0.065	0.079	-0.045	-0.090	-0.022	-0.018	-0.055
F-test: active-inactive=0	5.207	4.022	2.236	8.280	0.504	0.372	3.013
p-value, F-test	0.023	0.045	0.135	0.004	0.478	0.542	0.083
Mean dep. var	4.345	4.345	4.351	4.381	4.368	4.376	4.416
R-squared	0.042	-	-	-	-	-	-
Pseudo R-squared	-	0.012	0.079	0.080	0.084	0.080	0.085
No. of observations	818,711	818,711	783,077	698,349	747,291	759,287	649,429

Notes: T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects. Further groups of control variables described in section 3.2 are successively integrated. LSLAs being abandoned are excluded from all regressions. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Regardless of the model specification, the coefficients of the variables of main interest – 'active' and 'inactive' – have a negative sign and are statistically significant in all specifications. Due to the significant coefficient of the 'inactive' variable, a LSLA location choice bias cannot be rejected. The sign and significance level of the DiD estimator is therefore relevant.

In the lean models (columns (1) and (2)), the likelihood of women living in an area defined as either 'active' or 'inactive' has a negative and significant effect on the stated ideal number of children. The DiD estimator of the OLS regression in column (1) shows a statistically significant increase in fertility preference of 0.065 estimation points. In the lean Tobit model (column 2) the statistically significant increase accounts for 0.079 estimation points. The positive sign of the DiD estimator is in line with the expectation from the descriptive statistics.[24] Taking into account the LSLA location bias, LSLA operation has an increasing effect on the stated ideal number of children. Yet, the R^2 and pseudo R^2 of these models are rather low. Moreover, the DiD estimator is of rather small magnitude compared to the sample mean accounting for approximately four children.[25]

Including socio-demographic control variables into the regression (column (3)), the DiD estimator turns out to be negative albeit insignificant. Yet, the negative DiD estimator appears to be highly significant when wealth as a proxy for household income is included in the model (column (4)). Thereby, the increase in the size of the DiD estimator is mainly explained by a decrease in the magnitude of the negative 'inactive' coefficient. Although the magnitude of the 'inactive' coefficient decreases by more than half of its size when wealth is included in the regression, it is still significant at the 5 percent level. The LSLA location bias cannot be explained solely by wealth. Yet, wealth as a by-product determines the insignificant difference between the effect of living in 'active' areas on fertility desire compared to 'inactive' areas. Adding the proxy variables for female empowerment and maternal opportunity costs of time, the negative DiD estimator appears to remain insignificant.

Notably, controlling for wealth as a proxy of household income, the magnitude of the 'active' coefficient decreases as well. Compared to the decrease of the 'inactive' coefficient, the decrease of the 'active' coefficient is, however, less pronounced accounting only for a difference of 0.02 estimation points.[26] Including the proxies for maternal opportunity costs of time (column (5)) and female empowerment (column (6)), the coefficients of the 'active' variable are of greater magnitude than in the model including wealth. Consequently, incorporating these variables seem to be of less relevance in explaining the negative LSLA-fertility linkage. Moreover, these proxy variables even have a suppressing effect on the negative, 'inactive' coefficient compared to the model including socio-demographic control variables only (column (3)).

[24] As shown in the descriptive analysis, women living in 'active' areas have - on average - stated a higher ideal number of children than woman living in 'inactive' areas. However, the difference is insignificant and of rather small magnitude (see Table 1).

[25] The pseudo R^2 measures the goodness-of-fit of limited dependent variable models. Identical to the R^2 reported for OLS regressions, the measure takes on values between 0 and 1. However, it is based on log-likelihoods and therefore cannot be compared directly with the usual R^2 (Wooldridge, 2020).

[26] Substracting -0.136 (column (4)) from -0.156 (column (3)) yields a difference of 0.02 estimation points.

When all sets of variables are included (column (7)), the DiD estimator is negative and significant. While still being statistically significant, the magnitude of the 'active' coefficient – accounting for -0.106 estimation points – is smaller than in the Tobit model that exclusively comprises socio-demographic controls (column (4)). Consequently, including all proxies for the three defined mechanisms, the magnitude of the negative 'active' coefficient decreases. Nevertheless, still highly significant and negative, the impact of the proximity to 'active' LSLAs on fertility desire cannot be explained merely by these mediating variables.

Impact on number of children ever born

As described in section 4, for the dependent variable 'number of children ever born' the Tobit application to corner solution responses will be applied. Thereby, OLS and Tobit coefficients cannot be compared directly since – in opposite to the Tobit model for censoring – the interest relies in analysing the observed outcome y and not the latent variable y*. Therefore, all regression coefficients using the Tobit model for corner solution responses are reported as average marginal effects, which allows to compare the coefficients with the OLS results (Wooldridge, 2020).

In line with the previous findings, the effect of the 'active', independent variable on the number of children ever born has a highly significant, negative sign in all model specifications presented in Table 4. While also being negative, the average marginal effect of the 'inactive' variable turns out to be insignificant when household income proxied by the computed wealth score is included (column (4)). Hence, being located in an 'inactive' area decreases the number of children ever born merely as a by-product of household income proxied by wealth. Thereby, following the findings from the descriptive section, households located nearby LSLAs (regardless of the implementation status) are – on average – significantly wealthier than households living in 'control' areas. Yet, the differences of the means are rather small.

Employing the described DiD strategy, $(\beta_1 - \beta_2)$ is negatively signed and significant at one percent level throughout all model specifications except in the lean models as presented in columns (1) and (2).

Table 4: Effect of land deals on number of children ever born

	(1) OLS	(2) Tobit	(3) Tobit	(4) Tobit	(5) Tobit	(6) Tobit	(7) Tobit
Active 25 km	-0.254***	-0.262***	-0.133***	-0.127***	-0.128***	-0.134***	-0.120***
	(-8.30)	(-8.26)	(-5.76)	(-5.04)	(-5.76)	(-5.74)	(-4.78)
Inactive 25 km	-0.328***	-0.320***	-0.051***	-0.017	-0.043***	-0.053***	-0.011
	(-14.49)	(-13.62)	(-4.29)	(-1.24)	(-3.80)	(-4.35)	(-0.80)
Socio-Demographics	No	No	Yes	Yes	Yes	Yes	Yes
Household Income	No	No	No	Yes	No	No	Yes
Opportunity Costs	No	No	No	No	Yes	No	Yes
Female Empowerment	No	No	No	No	No	Yes	Yes
Difference in differences	0.074	0.057	-0.081	-0.110	-0.084	-0.082	-0.109
F-test: active-inactive=0	3.94	2.19	10.23	15.47	11.87	9.89	15.31
p-value, F-test	0.047	0.139	0.001	0.000	0.000	0.001	0.000
Mean dep. var	2.848	2.848	2.855	2.902	2.889	2.890	2.970
R-squared	0.004	-	-	-	-	-	-
Pseudo R-squared	-	0.000	0.301	0.303	0.305	0.299	0.304
No. of observations	890,237	890,237	853,203	764,497	800,003	810,702	696,786

Notes: Coefficients of the Tobit model are replaced by the average marginal effects. T- statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects. Further groups of control variables described in section 3.2 are successively integrated. LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01.*

Comparing the mediating impact of the three defined mechanisms, the magnitude of the negative, average marginal effect on fertility of living in an 'active' area compared to 'control' areas is weakest in column (4) and column (5) compared to column (6). Thus, proxies for household income and maternal opportunity costs of time play a (small) role in explaining the negative impact of women living in an 'active' area on the number of children ever born compared to women living in 'control' areas. Introducing the index measuring attitudes towards IPV against women as a proxy for female empowerment (column (6)), the average marginal effect appears to be even more strongly negative. Consequently, including female empowerment suppresses the negative LSLA-fertility association in this model.

Encompassing all relevant variables (column (7)), the 'inactive', independent variable turns out to be insignificant and thus the effect of the 'active' variable is not expected to suffer from a LSLA location selection bias. Similar to the previous results on fertility desire, most of the negative LSLA-fertility linkage is not explained by the three defined mechanisms. This can be seen as including all sets of variables, the likelihood of living in an 'active' area has a highly significant, average marginal effect on the ideal number of children still accounting for -0.120 estimation points.

In comparison to the findings on the desired family size (Table 3, column (7)), the pseudo R^2 of the specification including all variables accounts for 30.4 percent and is thus larger than the pseudo R^2 of the desired family size model accounting for 8.5 percent. Thus, this model measuring the impact on the number of children ever born provides a better goodness-of-fit than the model analysing the impact on the ideal number of children.

Impact on current fertility

Accounting for the binary nature of the dependent variable measuring current fertility, the Probit model is applied in addition to a simple OLS regression. Thereby, regression results using the Probit model are reported as average marginal effects. This allows to interpret coefficients with regard to a change in the predicted probability of the dependent variable (Wooldridge, 2020).

In Table 5, the results from the impact of LSLA proximity on the binary variable measuring fertility within the last 12 months are presented.

Table 5: Effect of land deals on current fertility

	(1) OLS	(2) Probit	(3) Probit	(4) Probit	(5) Probit	(6) Probit	(7) Probit
Active 25 km	-0.034***	-0.035***	-0.013***	-0.013***	-0.014***	-0.013***	-0.015***
	(-9.24)	(-8.91)	(-3.88)	(-3.51)	(-4.09)	(-3.86)	(-3.93)
Inactive 25 km	-0.039***	-0.041***	-0.009***	-0.006**	-0.009***	-0.008***	-0.007**
	(-13.62)	(-13.15)	(-3.77)	(-2.38)	(-3.63)	(-3.51)	(-2.55)
Socio-Demographics	No	No	Yes	Yes	Yes	Yes	Yes
Household Income	No	No	No	Yes	No	No	Yes
Opportunity Costs	No	No	No	No	Yes	No	Yes
Female Empowerment	No	No	No	No	No	Yes	Yes
Difference in differences	0.005	0.005	-0.004	-0.006	-0.005	-0.005	-0.008
F-test: active-inactive=0	1.320	-	-	-	-	-	-
Chi2-test: active-inactive=0		1.071	1.130	2.153	1.755	1.473	2.872
p-value, F/Chi2-test	0.251	0.301	0.288	0.142	0.185	0.225	0.090
Mean dep. var	0.264	0.264	0.265	0.266	0.267	0.269	0.270
R-squared	0.003	-	-	-	-	-	-
Pseudo R-squared	-	0.002	0.182	0.182	0.180	0.181	0.180
No. of observations	890,237	890,237	853,203	764,497	800,003	810,702	696,786

Notes: Coefficients of the Probit model are replaced by the average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects. Further groups of control variables described in section 3.2 are successively integrated. LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01.*

In harmony with the previous results, the average marginal effects of the independent variables are negative and highly significant throughout all specifications. Thus, a negative effect of LSLA proximity ('active' or 'inactive') on current fertility is expected.

Since the effect of living in an 'inactive' area is negative and highly significant throughout all model specifications, the existence of a selection bias by LSLAs location cannot be rejected in this model. Yet, identical to the previous findings, with different sets of independent variables added to the Tobit model (columns (3) – (6)), the DiD estimator turns out to be negative. Notably, the DiD estimator is only significant and of highest magnitude when all variables are included in the model (column (7)). In this model, the likelihood of having recently been pregnant or given birth of women living in 'active' compared to 'inactive' areas is on average 0.8 percentage points lower. Moreover, the average marginal effect of living in an 'active' compared to a 'control' areas accounts for a decrease of 1.5 percentage points.

Comparing the impact of the successive integration of the proxy variables (columns (4) to (6)), the negative association of living in an 'active' area on current fertility is weakest when wealth is introduced in the model (column (4)). This is consistent with the previous findings. The same holds for the 'inactive' average marginal effect. However, the reduction in magnitude is only minor. In harmony with the previous results, the DiD estimator is of greatest magnitude when wealth is introduced (column (4)) compared to the DiD estimator including the other sets of proxies (columns (5) and (6)). Yet, the effect of the negative DiD estimator accounting for 0.6 estimation points is rather small and the estimator insignificant.

Surprisingly, the average marginal effect of the 'active' variable on current fertility is more strongly negative when the proxies for maternal opportunity costs of time are included in the model (column (5)) compared to the model only including socio-demographic controls (column (3)). Consequently, proxies of maternal opportunity costs of time are likely to suppress the negative LSLA-fertility linkage. Following the descriptive analysis and the arguments from the conceptual framework, the significantly lower proportion of women currently working in targeted areas compared to control areas might explain the suppressing effect. While the magnitude is less strong, the proxy for female empowerment has as well a suppressing effect on the negative LSLA-fertility relationship.[27]

[27] Comparing the magnitude of the negative 'active' effect in column (6) and column (3).

To sum up, desired and realised fertility are expected to be lower for women surrounded by at least one 'active' LSLA in comparison to women living in (i) 'control' areas and (ii) women living nearby at least one 'inactive' LSLA. Thereby, the existence of a LSLA location bias cannot be rejected and is partly explained by wealth as a proxy for household income. A possible reason for a significantly negative 'inactive' coefficient in some model specifications might be that it captures pre-operating effects – e.g. jobs generated in the prospecting and investment phase of the LSLA establishment (Kotsadam and Tolonen, 2016). Based on the findings of the conceptual framework, employment creation was identified as a key pathway. Notably, including all sets of variables, the effect of the 'active' variable on fertility is still highly significant and negative, the impact of proximity to 'active' (i.e. operating) LSLAs on fertility cannot be explained merely through the three defined mediating mechanisms. However, household income proxied by wealth seems to contribute most to the negative LSLA-fertility association. The signs of the contribution of proxy variables of maternal opportunity costs of time and female empowerment, instead, remain ambiguous.

6.2 Heterogeneity analysis

In Table A12, the findings from the main model are tested by using exclusively DHS couples surveys. The dependent variable 'number of children ever born' no longer suffers from the corner solution problem. Thus, a simple OLS regression is applied. Socio-demographic control variables as described in subsection 3.2 are included.[28] The effects of the 'active' and 'inactive' independent variables are still negative, of greater magnitude and highly significant for all models. Yet, the DiD estimator in column (1) appears to be positive albeit it remains insignificant. The negative DiD estimator in column (2) is still significant and insignificant for the model measuring the impact on current fertility (column (3)). Adding husband's characteristics to the control variables and restricting the sample to couples surveys do not change the overall outcome.

Following Kubitza and Gehrke (2018), it is expected that the impact of LSLAs on fertility materialises over time and thus findings might vary when different time periods are considered. In Table

[28]The explanatory variables for the three mechanisms are not included in the subsection on heterogeneity and robustness because they explain – to some small extent – the negative association between LSLA and fertility. Therefore, the following results are compared with column (3) in each table of the main regressions. Thus, only socio-demographic controls are included.

A13, A14 and A15, the impact of 'active' LSLAs who have operated since (i) no more than five years (ii) between five and ten years and (iii) more than ten years is tested.[29] Surprisingly, the impact on the number of children ever born as well as current fertility is strongest – significant and negative – for 'active' LSLAs who have operated since less than five years (see column (1) of Tables A14 and A15).[30] Moreover, the effect of being located in an 'active' area turns out to be positive for LSLAs who have operated more than 10 years – albeit only weakly significant for the model regarding current fertility and insignificant for the one regarding children ever born, respectively. For the model measuring the impact on fertility preferences (Table A13), the magnitude of the negative coefficient of the 'active' variable only differs by 0.001 points between LSLAs who have operated less than five years and LSLAs who have operated over ten years (columns (1) and (3)). Moreover, in Table A13 all negative DiD estimators are insignificant and findings seem to be ambiguous. To sum up, against the expectation that a full impact on fertility can be measured when applying a long-period approach (Kubitza and Gehrke, 2018), the results show that reductions in reproductive outcomes are mostly linked to the proximity of LSLAs which are operating less than five years. Yet, it needs to be acknowledged that the proportion of women matched with LSLAs being 'active' more than ten years accounts for 0.20 percent whereby the proportion for LSLAs being 'active' less than five years accounts for 1.24 percent (see Table A9). The small sample size might be a reason for the decreased statistical power of proximity to LSLA who have operated more than ten years on realised fertility.

The main findings show that the presence of 'active' LSLAs has a negative effect on fertility desire and reproductive outcomes. Yet, the results do not provide any information about whether 'abandoned' LSLAs have an impact on fertility rates as they are excluded from the main regressions. In her paper exploring project failure of LSLAs globally, Nolte (2020) argues that project failure is likely to affect host regions adversely.[31] Environmental impacts such as deforestation, soil degradation, loss of biodiversity and water contamination might be a threat to people's livelihoods while positive effects such

[29] In each column, only women merged with 'active' LSLAs considering the chosen time span are included. Women merged with the respective other 'active' statuses are excluded. Thus, the sample size of the models differs depending on which time span is chosen (see Table A9).

[30] The DiD estimator is considered since in both tables the 'inactive' variable is significant.

[31] Project failure is defined as a LSLA that fails after it had started operation. Yet, Nolte (2020) includes project failure even before a contract is concluded or before it enters operation stages. Thus, it needs to be acknowledged that in this thesis only one type of potential project failure is analysed.

as contributions to rural development and changes in sectoral composition are missing. Table A16 shows that proximity to 'abandoned' LSLAs within a radius of 25 km has no significant effect on the three dependent variables whereby the sign of the effect appears to be positive. In section 7, possible explanations for a low representation of 'abandoned' LSLAs in the sample which might shape these results will be given. To put it in a nutshell, based on the data given, the negative LSLAs-fertility relationship does not seem to persist when LSLAs are abandoned.

6.3 Robustness

Several robustness tests have been performed to confirm the validity of the main findings. First, the challenge of Land Matrix data regarding consistent and accurate georeferencing (Messerli et al., 2014) will be addressed. A robustness test is performed which exclusively includes land deals with the highest level of spatial accuracy. This reduces the LSLA sample set from initially 367 to 48 locations. Table A17 in the appendix presents the results.[32] Measuring the impact on the ideal number of children (column (1)), the coefficient of the 'active' variable of main interest is still negative, of greater magnitude and highly significant. Thus, living close to at least one 'active' LSLA within 25 km – compared to the 'control' group – reduces the ideal number of children by 0.342. The DiD estimator is still insignificant. Yet, it turns out to be positive. Considering the impact on realised fertility, the average marginal effects of the 'active' variable remain negative albeit only significant in column (2). Moreover, the DiD estimators appear to be positive while being insignificant. Thus, this robustness check indicates that a lack of spatial accuracy in the locations of LSLAs could lead to biased results regarding the impact on current fertility. Moreover, the distinction of the impact of LSLAs on fertility by implementation status ('active' vs. 'inactive') is less clear.

In this thesis, it is assumed that the population is the same before and after the LSLA establishment. Therefore, migration might be a threat to the identification strategy. Following the approach of Kotsadam and Tolonen (2016), Table A18 presents the main regressions with the sample restricted to women who have never moved in their lives.[33] For each variable, the effects of the main independent

[32] The sample set is higher than in the main analysis and accounts for 903,259 women (the initial sample). This is because using a reduced LSLAs sample, all women could be assigned to one group – active, inactive or control.

[33] This reduces the sample size significantly including less than 230,000 observations in all model specifications.

’active’ variable are negative albeit insignificant for the model regarding current fertility (column (3)). It is noteworthy that for the fertility desire model the size of the coefficient of the ’active’ variable is more than twice as large as the coefficient in the main model (see Table 3, column (3)). Moreover, the negative DiD estimator is highly significant. To sum up, the hypothesis that female migration – particularly on current fertility – has an impact on the LSLAs-fertility linkage cannot be rejected.

Following Kotsadam and Tolonen (2016), robustness checks regarding distance were performed. Table A19 and A20 show that considering different definitions of the LSLAs footprint areas (15 km and 50 km cut-off distances), the results resemble the main results of the ’active’ variable both in terms of direction and significance. Overall, the choice of the distance cut-off seems not to bias the main findings. Yet, the negative DiD estimator appears to be significant in the model on fertility desire when a cut-off distance of 15 km is applied (see Table A19). Moreover, using a cut-off distance of 50 km, all three negative DiD estimators are highly significant. Another threat of the validity of the main findings could be that women living in areas far away from any LSLA may be too different for a meaningful comparison (Kotsadam and Tolonen, 2016). Therefore, in Table A21 the ’control’ group is restricted to only include women living less than 200 km away from any LSLA.[34] The results are consistent with the main findings. Thus, it seems to be reasonable to include, inter alia, women living more than 200 km away from any LSLA in the ’control’ group.

7 Caveats

Even though a heterogeneity analysis and various robustness tests have been performed, the following caveats need to be considered when interpreting the results of this thesis.

The validity of the findings hinges largely on the quality of the Land Matrix data and thus limitations regarding the Land Matrix dataset are introducing multiple biases. Due to conflicting information from several sources, the verification of basic data (such as location and implementing status) can be challenging. In many countries, procedures for decision-making on LSLAs are missing and negotiations do not take place in the public realm. Media attention often diminishes as soon as a deal is concluded, and thus its actual development remains uncertain. Additionally, decisions might be

[34] Thus, the sample is reduced by 215,639 observations corresponding to 24 percent of the sample (see Table A9).

changed without being communicated publicly. It occurs that intentions have been published, but no announcement was given when these deals were abandoned later. Thus, failed deals are likely to be under-represented in the dataset (Land Matrix, 2020a). These biases in the data collection provide an enormous threat to the identification strategy. For the DiD strategy applied, reliable information regarding LSLAs implementation status as well as its location is essential. Even though the data provided by the Land Matrix is imperfect, it is considered the most comprehensive, available database providing information about large-scale land deals on a global scale (Nolte, 2020; Nolte and Ostermeier, 2017).

Moreover, the classification of the three groups - active, inactive and abandoned - is based on year and not specific date information since the Land Matrix database only contains year information. Thus, all LSLAs that started operation in the year of the matched DHS interview are coded as ’inactive’. However, it is possible that a LSLA started operation in this particular year before the DHS interview took place and thus is erroneously classified as ’inactive’. In addition, the implementation status is missing for many LSLAs, although it is known that the contracts for these LSLAs have been signed. As these LSLAs are not included in the sample, this could pose another threat to the identification strategy. For example, some women living in ’control’ areas might actually be surrounded by at least one ’active’, ’inactive’ or ’abandoned’ LSLA that is not captured by the estimation strategy used.

A further limitation of the data used in this thesis is its cross-sectional nature. Using panel data instead would have allowed to evaluate changes over time regarding respondent’s individual characteristics and thus would be more suitable to analyse changes in fertility desires and outcomes evoked by LSLA proximity (Wooldridge, 2020).

Besides the data quality bias, it needs to be considered that this study provides only first hints on the relationship between LSLAs location and fertility on an aggregated level. Thereby, characteristics of LSLAs such as differences in deal size, intention and type of crop cultivated or promises of social investments are not accounted for. Following the arguments of Nolte and Ostermeier (2017), the type of crop being cultivated is of particular interest when analysing employment creation linkages and thus disentangling the effect by LSLAs characteristics might provide valuable findings.

Another shortcoming of this analysis is that it is based on assumptions of the simple fertility theory. The opportunity costs of time argument is analysed considering exclusively mother's time. More modern approaches in particular analysing partner's involvement in child-rearing or hiring nannies would alter the opportunity costs of time argument and introduce new monetary costs (Jones et al., 2008).[35] Moreover, as described in subsection 3.2, the proxies for – household income, maternal opportunity costs of time and female empowerment – are imperfect. For instance, introducing data on changes in land tenure and titling in future analysis might provide a valuable contribution to this research question. Land tenure and titling is expected to affect all three mechanisms and thus fertility through several pathways (Field, 2003). Moreover, analysing changes in female wages, type of payment as well as female employment on LSLAs might provide more information on the mediating effect of maternal opportunity costs of time.

Lastly, the empirical analysis is limited by analysing three mechanisms presented in the conceptual framework and the findings show that these mechanisms only explain a small part of the LSLA-fertility association. Yet, other mechanisms shaped by the proximity of LSLAs which might alter fertility preferences and outcomes are imaginable. For instance, the presence of child labour, returns to children's education, infrastructure development, or migration patterns might add additional value to this analysis (Kubitza and Gehrke, 2018). Following the argumentation in the introduction, studying migration patterns of young men could be of particular interest. Moreover, considering the findings from the telephone survey in Liberia suggesting promotion of rural development, examining social investments of LSLAs operators, such as investments in health facilities, could provide valuable insights on this topic. This might explain the significant lower average number of children who died in 'active' compared to 'inactive' and 'control' areas (see Table A10).[36]

[35] While husband's education level has been included as a control variable using DHS couples surveys, analysing the mediating effect of husband's education is not within the scope of this thesis.

[36] Yet, the magnitude of the differences is small and is likely to be linked to overall differences in fertility between these areas.

8 Conclusion

The aim of this thesis was to evaluate whether proximity to an operating large-scale land investment site has an impact on women's fertility and to identify the underlying mechanisms of this relationship. Three mechanisms – household income, maternal opportunity costs of time and female empowerment – likely to be altered by LSLA proximity and assumed to have an impact on fertility were identified. Considering the arguments of the conceptual framework, the impact of LSLAs on these mechanisms remains ambiguous and thus a positive or negative effect on fertility imaginable. Thereby, changes in employment seem to play a distinctive role.

Using a cut-off distance of 25 km, three groups of (i) women living close to at least one 'active' LSLA, (ii) women living close to at least one 'inactive' LSLA and (iii) women living in 'control' areas were defined. Based on descriptive statistics, an adverse impact of LSLA proximity ('active' or 'inactive') on fertility was assumed. Moreover, women living in 'active' and 'inactive' areas were identified to be – on average – wealthier, less women are currently working while more women working in the non-agriculture sector and have attained higher levels of education than women living in 'control' areas. Moreover, while using an imperfect proxy, women situated in 'active' areas are assumed to be more empowered.

Findings from the estimation results were in line with the descriptive statistics, as it can be concluded that geographical proximity to large-scale land acquisitions has a negative effect on women's fertility desires and reproductive outcomes. Introducing all independent and control variables into the model, the negative DiD estimator is significant regardless of which dependent variable is chosen. Thus, the empirical findings of this thesis confirm a negative LSLAs-fertility linkage as stated by the suggestive evidence of the telephone survey in Liberia. Yet, a LSLA location selection bias is likely and mostly explained by wealth. Following the discussion in the descriptive section, it remains unknown whether LSLAs are targeting more urban areas or whether due to the establishment of LSLAs rural development is promoted.

Regarding the three identified mechanisms, household income proxied by wealth seems to contribute most to the negative LSLA-fertility association. The sign of the contribution of proxy variables for maternal opportunity costs of time and female empowerment remains ambiguous. Yet, the results might differ if more suitable proxies would have been applied. Furthermore, since the mechanisms are highly linked to each other, interpreting them in isolation might not reflect their whole impact. Overall, it can be seen that the impact of proximity to 'active' (i.e. operating) LSLAs on fertility cannot be explained mainly through the three mediating mechanisms.

The heterogeneity analysis shows that analysing exclusively couples surveys, the main findings are confirmed. However, the hypothesis that the impact on fertility is highest when nearby LSLAs have been operating for a long time is rejected. Yet, the main findings are robust regarding several distance measures. Female migration and imprecise location information of land deals seem to play a distinctive role for measuring the impact on current fertility which is – due to the specific time horizon – the most accurate dependent variable in this thesis.

As stated in the introduction, a decline in fertility is directly associated with women's socio-economic well-being. Regarding LSLAs, vast literature provides evidence for negative effects of large-scale land investments on people's livelihoods surrounded by LSLAs and thus on women. Yet, this quantitative study provides first evidence for a positive linkage of LSLAs on women's socio-economic well-being whereby it is not expected that this linkage is universal. Considering the ambiguous findings from the theoretical discussion, LSLAs harbour the risk of worsen livelihoods of local people and women in particular. To minimise the risks, large-scale land investments need to be properly executed with special attention to the local context and gender dimensions. For instance, governments supporting land titling including both husband and wife might be favourable. Behrman et al. (2014) argue that this could reduce the risk of land expropriation and thus of losses in household income, as well as strengthen women's autonomy. Moreover, the risk of project failure and the long-term sustainability of LSLAs must be taken into account as the findings from the heterogeneity analysis show. To reduce the risk of project failure, Nolte (2020) proposes that host-country governments and international funders need to screen investment projects in agriculture more closely. Based on her analysis, Nolte (2020) recommends including domestic shareholders in investment projects and to avoid mega projects likely to fail as well as the usage of land formerly used by local communities. Moreover, adverse, long-term impacts on a variety of ecosystem functions have been documented for oil palm expansion in Indonesia (Kubitza and Gehrke, 2018). Consequently, in line with Kubitza and

Gehrke (2018), an assessment of several societal impacts of LSLAs should be conducted and weighted carefully against each other.

Overall, this thesis provides the first large-scale quantitative analysis that sheds light on the previously unexplored relationship between fertility and large-scale land investments on the African continent. Due to the limited scope of this thesis and data quality, the analysis contains various caveats. Further research is required to establish the link between LSLAs and fertility in surrounding areas more rigorously and to identify all underlying mechanisms.

References

Ainsworth, Martha, Kathleen Beegle, and Andrew Nyamete (1996). “The Impact of Women’s Schooling on Fertility and Contraceptive Use: A Study of Fourteen Sub-Saharan African Countries”. In: *The World Bank Economic Review* 10.1, pp. 85–122.

Ali, Daniel, Klaus Deininger, and Anthony Harris (2019). “Does Large Farm Establishment Create Benefits for Neighboring Smallholders? Evidence from Ethiopia”. In: *Land Economics* 95.1, pp. 71–90. DOI: 10.3368/le.95.1.71.

Annan, Jeannie et al. (2021). “Taking power: Women’s empowerment and household Well-being in Sub-Saharan Africa”. In: *World Development* 140, p. 105292. DOI: 10.1016/j.worlddev.2020.105292.

Baumert, Sophia et al. (2019). “Forgone opportunities of large-scale agricultural investment: A comparison of three models of soya production in Central Mozambique”. In: *World Development Perspectives* 16, p. 100145. DOI: 10.1016/j.wdp.2019.100145.

Becker, Gary S. and H. Gregg Lewis (1974). “Interaction between Quantity and Quality of Children”. In: *Economics of the Family: Marriage, Children, and Human Capital*. University of Chicago Press, pp. 81–90.

Behrman, Julia, Ruth Meinzen-Dick, and Agnes Quisumbing (2012). “The gender implications of large-scale land deals”. In: *Journal of Peasant Studies* 39.1, pp. 49–79. DOI: 10.1080/03066150.2011.652621.

Behrman, Julia A., Ruth Meinzen-Dick, and Agnes R. Quisumbing (2014). “An Interpretation of Large-Scale Land Deals Using Boserup’s Theories of Agricultural Intensification, Gender and Rural Development”. In: *Ester Boserup’s Legacy on Sustainability: Orientations for Contemporary Research*. Ed. by Marina Fischer-Kowalski, Anette Reenberg, Anke Schaffartzik, and Andreas Mayer. Human-Environment Interactions. Dordrecht: Springer Netherlands, pp. 189–202. DOI: 10.1007/978-94-017-8678-2_12.

Bongaarts, John (2010). “The causes of educational differences in fertility in Sub-Saharan Africa”. In: *Vienna Yearbook of Population Research* 8, pp. 31–50.

Bongaarts, John, Ann K. Blanc, and Katharine J. McCarthy (2019). "The links between women's employment and children at home: Variations in low- and middle-income countries by world region". In: *Population Studies* 73.2, pp. 149–163. DOI: 10.1080/00324728.2019.1581896.

Bongaarts, John and John Casterline (2013). "Fertility Transition: Is sub-Saharan Africa Different?" In: *Population and Development Review* 38.s1, pp. 153–168. DOI: https://doi.org/10.1111/j.1728-4457.2013.00557.x.

Bongaarts, John and John B Casterline (2018). "From Fertility Preferences to Reproductive Outcomes in the Developing World". In: *Population and Development Review* 44.4, pp. 793–809.

Campbell, Oona M. R., Wendy J. Graham, and Lancet Maternal Survival Series steering group (2006). "Strategies for reducing maternal mortality: getting on with what works". In: *Lancet (London, England)* 368.9543, pp. 1284–1299. DOI: 10.1016/S0140-6736(06)69381-1.

Chen, Lincoln C. et al. (1974). "Maternal Mortality in Rural Bangladesh". In: *Studies in Family Planning* 5.11, pp. 334–341. DOI: 10.2307/1965185.

Collier, Paul and Anthony Venables (2011). *Land Deals in Africa: pioneers and speculators*. OxCarre Working Paper 069. Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford. URL: https://econpapers.repec.org/paper/oxfoxcrwp/069.htm (visited on 03/11/2021).

Cotula, Lorenzo, International Institute for Environment and Development, Food and Agriculture Organization of the United Nations, and International Fund for Agricultural Development (2009). *Land grab or development opportunity?: agricultural investment and international land deals in Africa*. London/Rome: IIED/FAO/IFAD. ISBN: 9781843697411.

Croft, Trevor N., Aileen M. J. Marshall, and Courtney K. Allen (2018). *Guide to DHS Statistics*. Tech. rep. Rockville, Maryland, USA: ICF.

Deininger, Klaus and Derek Byerlee (2012). "The Rise of Large Farms in Land Abundant Countries: Do They Have a Future?" In: *World Development* 40.4, pp. 701–714. DOI: 10.1016/j.worlddev.2011.04.030.

Deininger, Klaus and Fang Xia (2016). "Quantifying Spillover Effects from Large Land-based Investment: The Case of Mozambique". In: *World Development* 87, pp. 227–241. DOI: 10.1016/j.worlddev.2016.06.016.

Dodd, My-Lan et al. (2018). *Women's Land Rights in Liberia in Law, Practice, and Future Reforms: LGSA Women's Land Rights Study*. Tech. rep. United States Agency of International Development (USAID). URL: https://www.land-links.org/wp-content/uploads/2018/03/USAID_Land_Tenure_LGSA_WLR_Study_Mar-16-2018.pdf (visited on 03/20/2021).

Doepke, Matthias (Mar. 2015). "GARY BECKER ON THE QUANTITY AND QUALITY OF CHILDREN". In: *Journal of Demographic Economics* 81.1, pp. 59–66. DOI: 10.1017/dem.2014.8.

Dorward, Andrew (2013). "Agricultural labour productivity, food prices and sustainable development impacts and indicators". In: *Food Policy* 39, pp. 40–50. DOI: 10.1016/j.foodpol.2012.12.003.

Easterlin, Richard A. (1975). "An Economic Framework for Fertility Analysis". In: *Studies in Family Planning* 6.3, pp. 54–63. DOI: 10.2307/1964934.

Elmhirst, R., B. Sijapati Basnett, M. Siscawati, and D. Ekowati (2017). *Gender Issues in Large-scale Land Acquisition: Insights from Oil Palm in Indonesia*. Tech. rep. Washington, D.C.: Center for International Forestry Research (CIFOR). URL: https://www.cifor.org/knowledge/publication/6774/ (visited on 02/09/2021).

Ezeh, Alex, Frances Kissling, and Peter Singer (2020). "Why sub-Saharan Africa might exceed its projected population size by 2100". In: *The Lancet* 396.10258, pp. 1131–1133.

Field, Erica (2003). *Fertility Responses to Land Titling: The Roles of Ownership Security and the Distribution of Household Assets*. URL: https://scholar.harvard.edu/files/field/files/field_fertility_05.pdf (visited on 03/20/2021).

Food and Agriculture Organization (FAO) (2016). *Decent rural employment, productivity effects and poverty reduction in sub-Saharan Africa*. Technical Papers Series 5. Rome: Food and Agriculture Organization of the United Nations. URL: http://www.fao.org/reduce-rural-poverty/resources/resources-detail/ar/c/1029952/ (visited on 02/09/2021).

Gyapong, Adwoa Yeboah (2020). “How and why large scale agricultural land investments do not create long-term employment benefits: A critique of the ‘state’ of labour regulations in Ghana”. In: *Land Use Policy* 95, p. 104651. DOI: 10.1016/j.landusepol.2020.104651.

Haq, Iqramul, Saifullah Sakib, and Ashis Talukder (2017). “Sociodemographic Factors on Contraceptive Use among Ever-Married Women of Reproductive Age: Evidence from Three Demographic and Health Surveys in Bangladesh”. In: *Medical Sciences* 5.4, p. 31. DOI: 10.3390/medsci5040031.

Heath, Rachel and A. Mushfiq Mobarak (2015). “Manufacturing growth and the lives of Bangladeshi women”. In: *Journal of Development Economics* 115.C, pp. 1–15.

Herrmann, Raoul and Ulrike Grote (2015). “Large-scale Agro-Industrial Investments and Rural Poverty: Evidence from Sugarcane in Malawi”. In: *Journal of African Economies* 24.5, pp. 645–676. DOI: 10.1093/jae/ejv015.

Herrmann, Raoul T. (2017). “Large-Scale Agricultural Investments and Smallholder Welfare: A Comparison of Wage Labor and Outgrower Channels in Tanzania”. In: *World Development* 90, pp. 294–310. DOI: 10.1016/j.worlddev.2016.10.007.

Holden, Stein T. and Keijiro Otsuka (2014). “The roles of land tenure reforms and land markets in the context of population growth and land use intensification in Africa”. In: *Food Policy* 48, pp. 88–97. DOI: 10.1016/j.foodpol.2014.03.005.

ICF (2021). *The DHS Program - GPS Data Collection*. URL: https://dhsprogram.com/methodology/gps-data-collection.cfm (visited on 02/20/2021).

Jones, Larry, Alice Schoonbroodt, and Michèle Tertilt (2008). *Fertility Theories: Can They Explain the Negative Fertility-Income Relationship?* Tech. rep. w14266. Cambridge, MA: National Bureau of Economic Research, w14266. DOI: 10.3386/w14266.

Kirk, Dudley (1996). “Demographic Transition Theory”. In: *Population Studies* 50.3, pp. 361–387. DOI: 10.1080/0032472031000149536.

Kiser, Humayun and Md. Alamgir Hossain (2018). “Estimation of number of ever born children using zero truncated count model: evidence from Bangladesh Demographic and Health Survey”. In: *Health Information Science and Systems* 7.1. DOI: 10.1007/s13755-018-0064-y.

Kishor, Sunita and Lekha Subaiya (2008). *Understanding women's empowerment: a comparative analysis of demographic and health surveys (DHS) data*. DHS Comparative Reports 20. Calverton, Maryland, USA: Macro International. URL: https://dhsprogram.com/publications/publication-cr20-comparative-reports.cfm (visited on 03/12/2021).

Kleemann, Linda and Rainer Thiele (2015). "Rural welfare implications of large-scale land acquisitions in Africa: A theoretical framework". In: *Economic Modelling* 51, pp. 269–279. DOI: 10.1016/j.econmod.2015.08.016.

Knutsen, Carl Henrik, Andreas Kotsadam, Eivind Hammersmark Olsen, and Tore Wig (2017). "Mining and Local Corruption in Africa". In: *American Journal of Political Science* 61.2, pp. 320–334. DOI: https://doi.org/10.1111/ajps.12268.

Kotsadam, Andreas and Anja Tolonen (2016). "African Mining, Gender, and Local Employment". In: *World Development* 83, pp. 325–339. DOI: 10.1016/j.worlddev.2016.01.007.

Kotsadam, Andreas et al. (2018). "Development aid and infant mortality. Micro-level evidence from Nigeria". In: *World Development* 105, pp. 59–69. DOI: 10.1016/j.worlddev.2017.12.022.

Kubitza, Christoph and Esther Gehrke (2018). *Why does a labor-saving technology decrease fertility rates? Evidence from the oil palm boom in Indonesia*. Tech. rep. 22. University of Goettingen, Collaborative Research Centre 990 "EFForTS, Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems (Sumatra, Indonesia)". URL: https://ideas.repec.org/p/zbw/crc990/22.html (visited on 11/16/2020).

Land Matrix (2020a). *Frequently Asked Questions (FAQs)*. URL: https://landmatrix.org/faq/ (visited on 11/16/2020).

— (2020b). *Region - Africa*. URL: https://landmatrix.org/region/africa/ (visited on 11/16/2020).

Lay, Jann and Kerstin Nolte (2018). "Determinants of foreign land acquisitions in low- and middle-income countries". In: *Journal of Economic Geography* 18.1, pp. 59–86. DOI: 10.1093/jeg/lbx011.

Lay, Jann, Kerstin Nolte, and Kacana Sipangule (2018). *Large-Scale Farms and Smallholders: Evidence from Zambia*. Working Paper 310. GIGA Working Papers. URL: https://www.econstor.eu/handle/10419/175430 (visited on 02/11/2021).

Li, Tania Murray (2011). "Centering labor in the land grab debate". In: *The Journal of Peasant Studies* 38.2, pp. 281–298. DOI: 10.1080/03066150.2011.559009.

Maertens, Miet, Liesbeth Colen, and Johan F. M. Swinnen (2011). "Globalisation and poverty in Senegal: a worst case scenario?" In: *European Review of Agricultural Economics* 38.1, pp. 31–54. DOI: 10.1093/erae/jbq053.

Messerli, Peter et al. (2014). "The geography of large-scale land acquisitions: Analysing socio-ecological patterns of target contexts in the global South". In: *Applied Geography* 53, pp. 449–459. DOI: 10.1016/j.apgeog.2014.07.005.

Miller, Grant (2010). "Contraception as Development? New Evidence from Family Planning in Colombia". In: *The Economic Journal* 120.545, pp. 709–736. DOI: https://doi.org/10.1111/j.1468-0297.2009.02306.x.

Nolte, Kerstin (2020). "Doomed to fail? Why some land-based investment projects fail". In: *Applied Geography* 122, p. 102268. DOI: 10.1016/j.apgeog.2020.102268.

Nolte, Kerstin, Markus Giger, and Wytske Chamberlain (2016). *International Land Deals for Agriculture. Fresh insights from the Land Matrix: Analytical Report II*. Tech. rep. Bern, Montpellier, Hamburg, Pretoria: Centre for Development, Environment, University of Bern; Centre de coopération internationale en recherche agronomique pour le développement; German Institute of Global, and Area Studies; University of Pretoria. URL: https://boris.unibe.ch/85304/ (visited on 02/10/2021).

Nolte, Kerstin and Martin Ostermeier (2017). "Labour Market Effects of Large-Scale Agricultural Investment: Conceptual Considerations and Estimated Employment Effects". In: *World Development* 98.C, pp. 430–446.

Rutstein, Shea O. and Kiersten Johnson (2004). *The DHS Wealth Index*. DHS Comparative Reports 6. Calverton, Maryland USA: ORC Macro. URL: https://dhsprogram.com/publications/publication-cr6-comparative-reports.cfm (visited on 02/05/2021).

Ryan, Caitlin (2018). “Large-scale land deals in Sierra Leone at the intersection of gender and lineage”. In: *Third World Quarterly* 39.1, pp. 189–206. DOI: 10.1080/01436597.2017.1350099.

Schoneveld, George, Laura German, and Eric Nutakor (2011). “Land-based Investments for Rural Development? A Grounded Analysis of the Local Impacts of Biofuel Feedstock Plantations in Ghana”. In: *Ecology and Society* 16.4. DOI: 10.5751/ES-04424-160410.

Schoneveld, George C. (July 2017). “Host country governance and the African land rush: 7 reasons why large-scale farmland investments fail to contribute to sustainable development”. In: *Geoforum* 83, pp. 119–132. DOI: 10.1016/j.geoforum.2016.12.007. URL: https://www.sciencedirect.com/science/article/pii/S0016718516301671 (visited on 03/10/2021).

Schuler, Sidney Ruth, Rachel Lenzi, and Kathryn M. Yount (2011). “Justification of Intimate Partner Violence in Rural Bangladesh: What Survey Questions Fail to Capture”. In: *Studies in family planning* 42.1, pp. 21–28.

Sen, Amartya (1999). *Development as Freedom*. New York: Alfred A. Knopf, Inc.

Sen, G. and S. Batliwala (2000). “Empowering women for reproductive rights.” In: *Women’s empowerment and demographic processes*. New York, New York: Oxford University Press, pp. 15–36.

Upadhyay, Ushma D. and Deborah Karasek (2012). “Women’s Empowerment and Ideal Family Size: An Examination of DHS Empowerment Measures In Sub-Saharan Africa”. In: *International Perspectives on Sexual and Reproductive Health* 38.2, pp. 78–89.

Upadhyay, Ushma D. et al. (Aug. 2014). “Women’s empowerment and fertility: A review of the literature”. In: *Social Science & Medicine* 115, pp. 111–120. DOI: 10.1016/j.socscimed.2014.06.014.

Van den Broeck, Goedele, Johan Swinnen, and Miet Maertens (2017). “Global value chains, large-scale farming, and poverty: Long-term effects in Senegal”. In: *Food Policy* 66, pp. 97–107. DOI: 10.1016/j.foodpol.2016.12.003.

Wooldridge, Jeffrey M. (2020). *Introductory econometrics: a modern approach*. Seventh edition. Boston, MA: Cengage Learning. ISBN: 978-1-337-55886-0.

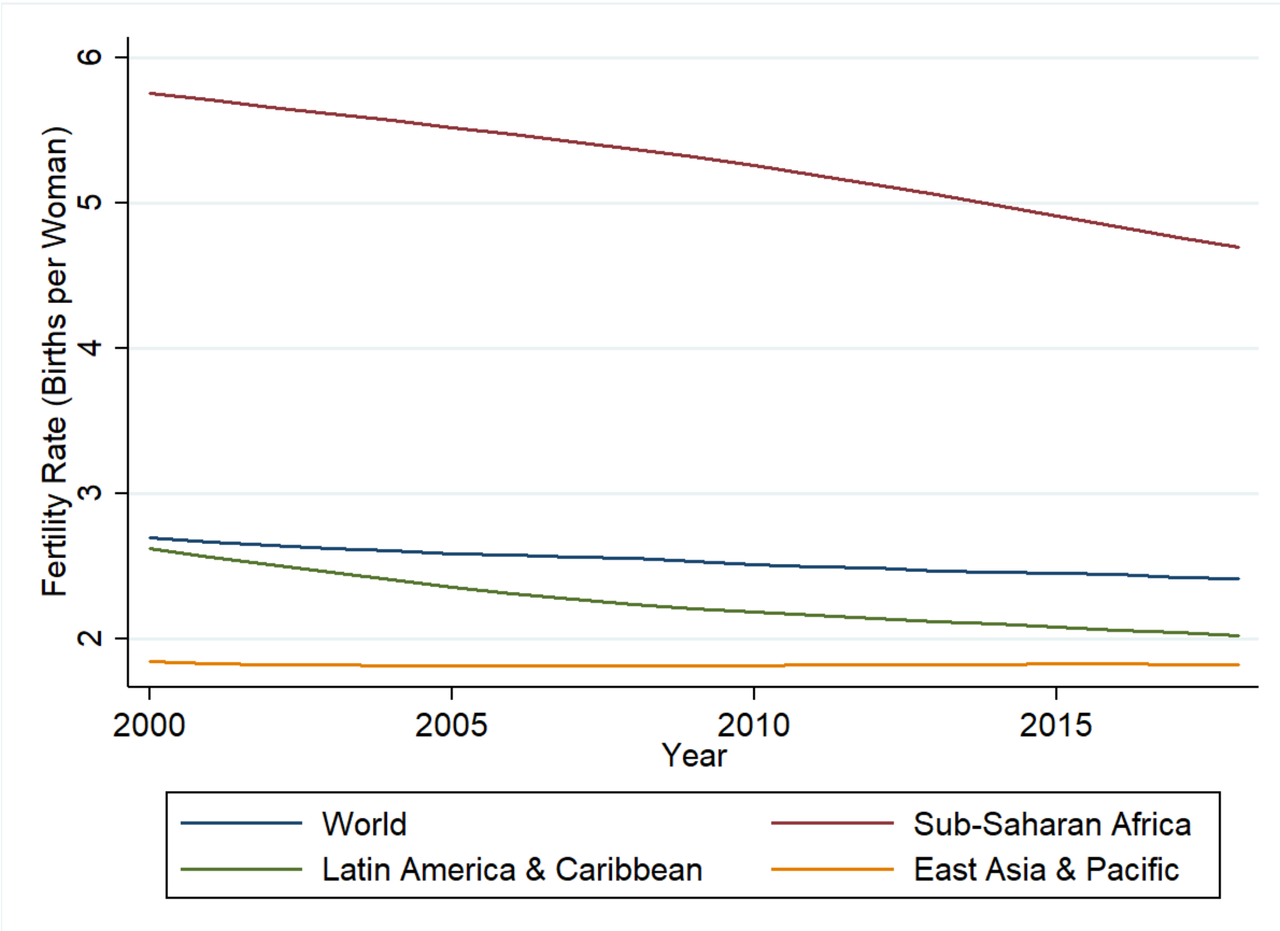
World Bank (2021a). *Fertility rate, total (births per woman) - Sub-Saharan Africa | Data*. URL: https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=ZG&most_recent_value_desc=false (visited on 02/02/2021).

— (2021b). *Population growth (annual %) - Sub-Saharan Africa | Data*. URL: https://data.worldbank.org/indicator/SP.POP.GROW?locations=ZG (visited on 02/12/2021).

A Appendix

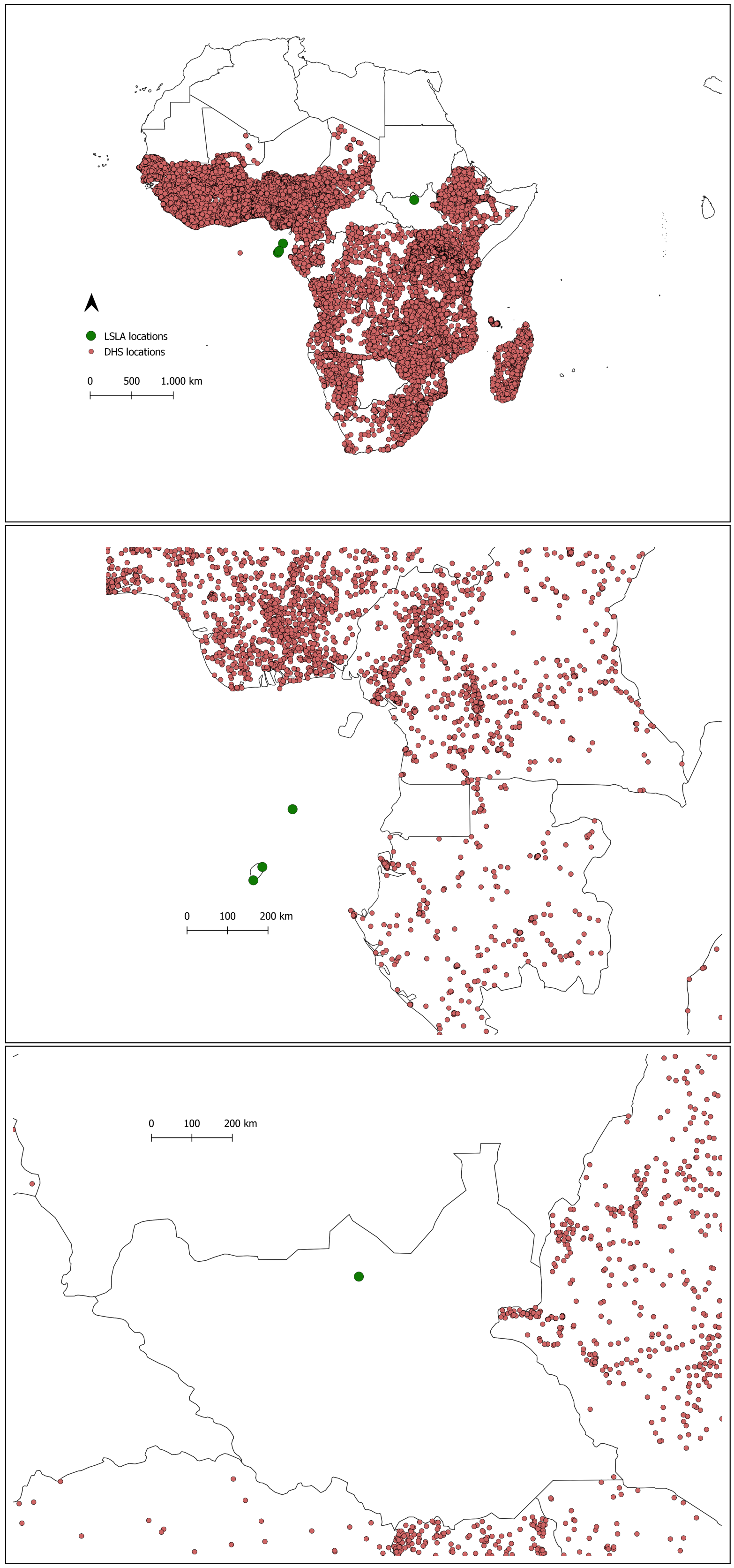
A.1 Descriptive statistics

Figure A2: Regional trends in fertility



(Own visualisation based on data from World Bank (2021a)

Figure A3: Location of land deals - Not matched with DHS clusters



(Own visualisation based on DHS and Land Matrix data)

Table A6: Comparison of fertility rates (births per woman)

	Fertility rates in 2016	Difference to SSA mean
Sub-Saharan Africa	4.84	/
Republic of the Congo	4.54	-0.30
Ethiopia	4.46	-0.38
Ghana	3.98	-0.85
Liberia	4.46	-0.38
Madagascar	4.18	-0.65
Mozambique	4.99	0.15
Sudan	4.53	-0.31
Sierra Leone	4.46	-0.38
South Sudan	4.86	0.02
Zambia	4.81	-0.02

Notes: Own visualisation and computation based on data provided by the World Bank, 2021a. Instead of using most recent data, figures from 2016 were chosen because the ten countries were selected based on Nolte et al. (2016)'s report.

Table A7: Distribution of land deals by host country

Countries	No of LSLA locations
Angola	9
Benin	1
Burkina Faso	1
Central African Republic	2
Côte d'Ivoire	4
Cameroon	18
Democratic Republic of Congo	13
Congo	5
Ethiopia	46
Gabon	7
Ghana	43
Guinea	3
Kenya	7
Liberia	9
Madagascar	9
Mali	4
Mozambique	46
Mauritania	1
Malawi	5
Namibia	3
Nigeria	24
Rwanda	1
Senegal	15
Sierra Leone	17
South-Sudan	2
Eswatini	1
Tanzania	18
Uganda	15
South Africa	2
Zambia	28
Zimbabawe	4
Total	363

Note: The number of locations of land deals per host country used in the final sample is shown.

Table A8: DHS sample disaggregated by country and interview year

	2000	2001	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
Angola	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7266	7113	0	0	0	14379
Benin	0	6219	0	0	0	0	0	0	0	0	1909	14690	0	0	0	0	7387	8541	0	38746
Burkina Faso	0	0	12429	0	0	0	0	0	0	16981	0	0	0	0	0	0	0	0	0	29410
Burundi	0	0	0	0	0	0	0	0	0	8151	1238	0	0	0	0	10012	7257	0	0	26658
Cameroon	0	0	0	10656	0	0	0	0	0	0	15426	0	0	0	0	0	0	13449	78	39609
Chad	0	0	0	0	0	0	0	0	0	0	0	0	0	5886	11833	0	0	0	0	17719
Comoros	0	0	0	0	0	0	0	0	0	0	0	5329	0	0	0	0	0	0	0	5329
Congo Democratic Republic	0	0	0	0	0	0	9995	0	0	0	0	0	14646	4181	0	0	0	0	0	28822
Cote d'Ivoire	0	0	0	0	0	0	0	0	0	0	1357	8703	0	0	0	0	0	0	0	10060
Eswatini	0	0	0	0	0	3813	1174	0	0	0	0	0	0	0	0	0	0	0	0	4987
Ethiopia	0	0	16515	0	0	0	0	15683	0	0	0	0	0	0	0	0	0	0	0	32198
Gabon	0	0	0	0	0	0	0	0	0	0	0	8422	0	0	0	0	0	0	0	8422
Ghana	0	0	5691	0	0	0	0	4916	0	0	0	0	0	9396	0	0	0	0	0	20003
Guinea	0	0	0	0	7954	0	0	0	0	0	0	9142	0	0	0	0	0	10874	0	27970
Kenya	0	0	8195	0	0	0	0	3924	4520	0	0	0	0	31079	0	0	0	0	0	47718
Lesotho	0	0	0	0	0	0	0	0	5225	991	0	0	0	5379	0	0	0	0	0	11595
Liberia	0	0	0	0	0	515	6577	0	0	0	0	0	9239	0	0	0	0	0	0	16331
Madagascar	0	0	0	0	0	0	0	3952	12007	0	0	0	0	0	0	0	0	0	0	15959
Malawi	13220	0	0	0	0	0	0	0	0	23020	0	0	0	0	16606	7956	0	0	0	60802
Mali	0	11519	0	0	0	0	0	0	0	0	0	5558	3720	0	0	0	0	10519	0	31316
Mozambique	0	0	0	0	0	0	0	0	0	0	13704	0	0	0	0	0	0	0	0	13704
Namibia	6755	0	0	0	0	2742	7062	0	0	0	0	0	9176	0	0	0	0	0	0	25735
Nigeria	0	0	7616	0	0	0	0	33346	0	0	0	0	38948	0	0	0	0	41821	0	121731
Rwanda	0	0	0	0	0	0	615	6698	0	7364	6307	0	0	4893	8604	0	0	0	0	34481
Senegal	0	0	0	0	14602	0	0	0	0	6437	9251	0	0	0	0	8865	0	0	0	39155
Sierra Leone	0	0	0	0	0	0	0	7374	0	0	0	0	16658	0	0	0	0	0	0	24032
South Africa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7343	0	0	0	7343
Tanzania	0	0	0	0	0	0	0	0	1038	9101	0	0	0	0	10206	3060	0	0	0	23405
Togo	0	0	0	0	0	0	0	0	0	0	0	0	3717	5763	0	0	0	0	0	9480
Uganda	3601	3645	0	0	0	8531	0	0	0	0	8674	0	0	0	0	18506	0	0	0	42957
Zambia	0	0	0	0	0	0	7146	0	0	0	0	0	9916	6495	0	0	0	13327	356	37240
Zimbabwe	0	0	0	0	6682	2065	0	0	0	4931	4147	0	0	0	9955	0	0	0	0	27780
Total	23576	21383	50446	10656	29238	17666	32569	75893	22790	76976	62013	51844	106020	73072	64470	62855	14644	98531	434	895076

(Own visualisation of DHS data)

Table A9: Descriptive statistics – Whole sample

Variables	Mean/ Proportion	Min	Max	Observations
Dependent variables				
Ideal number of children	4.34	0	6	823,440
Number of children ever born	2.85	0	18	895,076
Current fertility (%)	26.37	0	100	895,076
LSLA variables				
Distance closest LSLA (km)	154.74	0	1,651	895,076
Active 15km (%)	1.30	0	100	895,076
Inactive 15km (%)	3.40	0	100	895,076
Abandoned 15km (%)	0.33	0	100	895,076
Active 25km (%)	2.58	0	100	895,076
Less than 5 years (%)	1.24	0	100	882,863
Between 5 and 10 years (%)	0.99	0	100	880,640
More than 10 years (%)	0.20	0	100	873,681
Inactive 25km (%)	5.95	0	100	895,076
Abandoned 25km (%)	0.54	0	100	895,076
Active 50km (%)	6.62	0	100	895,076
Inactive 50km (%)	12.62	0	100	895,076
Abandoned 50km (%)	1.16	0	100	895,076
Active 200km (%)	43.47	0	100	895,076
Inactive 200km (%)	30.01	0	100	895,076
Abandoned 200km (%)	2.43	0	100	895,076
Not matched with any LSLA within 200km (%)	24.09	0	100	895,076
Mechanisms				
Wealth score	2.69	0	10	804,372
Highest educational level (%):				895,036
No education	32.74	0	100	
Primary	33.84	0	100	
Secondary	29.09	0	100	
Higher	4.33	0	100	
Currently working (%)	59.12	0	100	869,609
Non-agricultural work (%)	34.21	0	100	843,088
Not condones IPV (%)	51.56	0	100	851,803
Socio-demographics				
Age (single years)	28.43	15	49	895,076
Marital status (%):				895,066
Never in union	27.08	0	100	
Married	54.87	0	100	
Living with partner	9.43	0	100	
Widowed	2.92	0	100	
Divorced	2.22	0	100	

Variables	Mean/ Proportion	Min	Max	Observations
Separated	3.49	0	100	
Age First Cohabitation (%):				648,039
Younger than 15	16.06	0	100	
15-19	53.10	0	100	
20-24	22.23	0	100	
25-29	6.35	0	100	
30-39	2.11	0	100	
40 and older	0.16	0	100	
Rural area (%)	63.88	0	100	895,076
Household size	6.76	1	74	895,076
Female household head (%)	27.04	0	100	895,073
Religion (%):				858,010
No religion	1.41	0	100	
Christian	64.4 6	0	100	
Muslim	31.23	0	100	
Traditional	1.64	0	100	
Other	1.26	0	100	
Use of modern methods (%)	19.49	0	100	895,076
Husband's educational level (%):				597,092
No education	35.30	0	100	
Primary	29.90	0	100	
Secondary	27.43	0	100	
Higher	7.38	0	100	
Husband's ideal number of children	4.83	0	6	171,468
Number of living children	2.45	0	16	895,076
Number of children died	0.39	0	15	895,076
Migration non-mover (%)	47.77	0	100	511,019

Notes: Summary statistics are based on own calculations and DHS data. Statistics are reported in percentages where stated otherwise sample averages are given. Moreover, the minimum and maximum value and number of observations of each variable are given.

Table A10: Balance Check – Control variables

Variable	(1) Mean Control	(2) Mean Active	(3) Mean Inactive	(4) Active vs Control	(5) Inactive vs Control	(6) Active vs Inactive
Rural area	0.657	0.510	0.443	0.000***	0.000***	0.002***
Age (in years)	28.446	28.466	28.235	0.787	0.000***	0.009***
Marital status:						
Never in union	0.266	0.306	0.319	0.000***	0.000***	0.034**
Married	0.555	0.472	0.484	0.000***	0.000***	0.199
Living with partner	0.093	0.123	0.103	0.000***	0.005***	0.003***
Widowed	0.029	0.032	0.030	0.015**	0.276	0.132
Divorced	0.022	0.026	0.023	0.002***	0.198	0.052*
Separated	0.034	0.041	0.041	0.000***	0.000***	0.928
Age first cohabitation:						
Younger than 15	0.163	0.125	0.136	0.000***	0.000***	0.029**
15-19	0.532	0.511	0.521	0.000***	0.003***	0.162
20-24	0.220	0.254	0.242	0.000***	0.000***	0.033**
25-29	0.062	0.079	0.074	0.000***	0.000***	0.185
30-39	0.021	0.028	0.025	0.000***	0.000***	0.060*
40 and older	0.002	0.002	0.002	0.026**	0.042**	0.447
Religion:						
None	0.014	0.009	0.016	0.000***	0.414	0.001***
Christian	0.643	0.648	0.643	0.773	0.993	0.800
Muslim	0.314	0.317	0.311	0.853	0.804	0.769
Traditional	0.018	0.002	0.004	0.000***	0.000***	0.006***
Other	0.011	0.024	0.027	0.000***	0.000***	0.448
Female-headed household	0.265	0.319	0.327	0.000***	0.000***	0.250
Household size	6.730	7.293	7.112	0.000***	0.000***	0.312
Use of modern contraceptives method	0.192	0.231	0.209	0.000***	0.000***	0.000***
Husband's ideal number of children	4.845	4.639	4.667	0.000***	0.000***	0.533
Husband's educational level:						
No education	0.361	0.252	0.282	0.000***	0.000***	0.052*
Primary	0.301	0.303	0.265	0.857	0.000***	0.001***
Secondary	0.266	0.357	0.352	0.000***	0.000***	0.657
Higher	0.071	0.088	0.101	0.000***	0.000***	0.031**
Number of living children	2.474	2.280	2.192	0.000***	0.000***	0.003***
Number of children died	0.402	0.296	0.344	0.000***	0.000***	0.000***
Migration - Never moved	0.483	0.351	0.454	0.000***	0.000***	0.000***
Observations	813,879	23,130	53,228	837,009	867,107	76,358

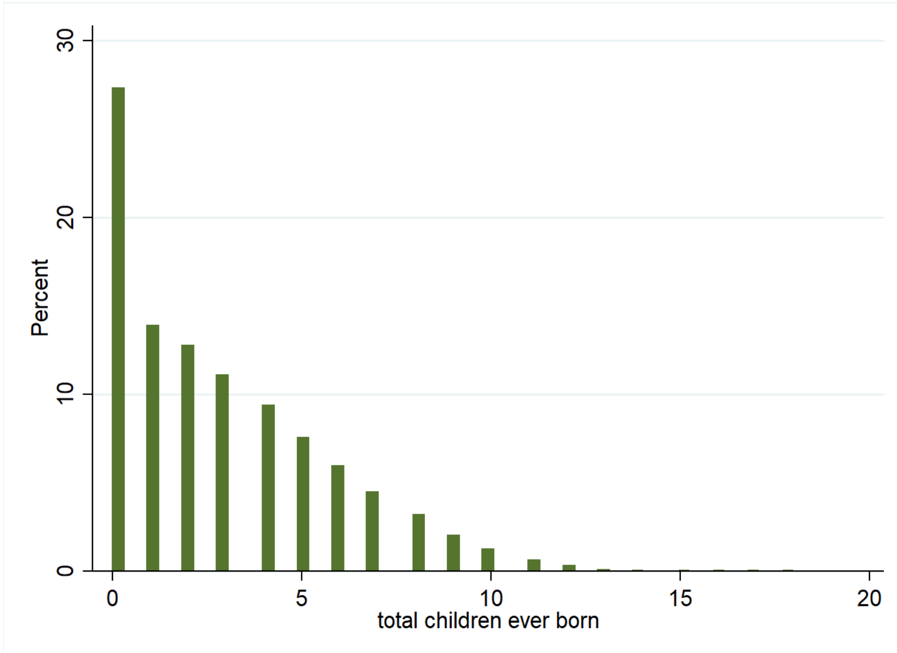
Notes: Columns (1) to (3) report the mean values disaggregated by the three groups – active, inactive and control – within 25km radius. Columns (4) to (6) report the p-values of the differences between means of the three groups. For categorical variables, for each category a dichotomous variable was created. Women matched with abandoned LSLAs within 25 km are excluded. Robust standard errors are clustered at the level of DHS primary sampling unit. **** p<0.01, ** p<0.05, * p<0.1.*

Table A11: Descriptive statistics – Telephone survey in Liberia

	Statistics	Observations
Women Questionnaire		
Reasons for having another child:		44
old-age support	4	
child help working/ contribute to household income	1	
preferred gender of child	15	
previous child died	1	
desire to have a big family	21	
other reasons	2	
Reasons for no additional child:		66
no financial means	19	
rather prefer to work	3	
prefer investing in children's education	41	
other reasons	3	
Change of child demand since plantation establishment:		84
increased a lot	2	
increased a little	10	
stayed the same	40	
decreased a little	29	
decreased a lot	3	
Town Chief Questionnaire		
Land access through permission of town chief or elders	30	30
Women working on plantation	19	30
Women help husbands on plantation	17	30
Increase of number of women working on plantation over time	15	19
Increase of number of women working in other jobs over time	15	26
Change of maternal opportunity costs of time over last couple years:		
Whole sample :		30
much more	11	
a little more	3	
unchanged	0	
a little less	2	
much less	14	
Restricted to chiefs reporting that more women work over time :		15
much more	0	
a little more	1	
unchanged	0	
a little less	2	
much less	12	
Change in rural development (modern buildings, more clinics etc.)	25	30

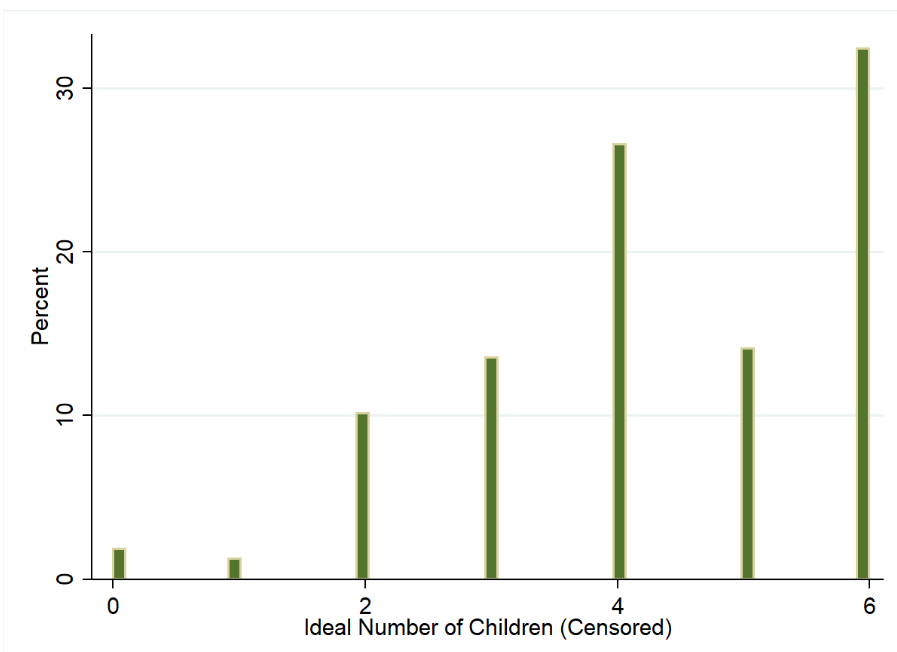
Notes: Statistics are based on own calculations and reported in total numbers. The respective number of observations per question is given and differs between several variables since some questions were follow-up questions or not mandatory. More details on the survey design are provided in section A.3. The reference period over temporal changes is considered to be the time since the establishment of the plantation.

Figure A4: Histogram – Number of children ever born



(Own visualisation based on DHS data)

Figure A5: Histogram – Ideal number of children



(Own visualisation based on DHS data)

A.2 Heterogeneity and robustness statistics

Table A12: Heterogeneity – Couples surveys

	(1) Tobit - Ideal No Children	(2) OLS - Children Born	(3) Probit - Current Fertility
Active 25 km	-0.171***	-0.134***	-0.022***
	(-4.91)	(-4.31)	(-2.89)
Inactive 25 km	-0.187***	-0.069***	-0.015***
	(-6.23)	(-3.44)	(-2.71)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	0.017	-0.065	-0.007
F-test: active-inactive=0	0.141	3.201	-
Chi2-test: active-inactive=0	-	-	0.648
p-value, F/Chi2-test	0.708	0.074	0.421
Mean dep. var	4.604	3.703	0.382
R-squared	-	0.706	-
Pseudo R-squared	0.094	-	0.128
No. of observations	152,496	161,594	161,594

Notes: In column (3), the coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as definded in subsection 3.2. The sample is reduced to couples only. LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01*.

Table A13: Heterogeneity – Time spans (Ideal number of children)

	(1) Tobit - Ideal Number of Children	(2) Tobit - Ideal Number of Children	(3) Tobit - Ideal Number of Children
Less than 5 years	-0.161***		
	(-4.42)		
Between 5 and 10 years		-0.122***	
		(-2.98)	
More than 10 years			-0.162**
			(-2.44)
Inactive 25 km	-0.111***	-0.109***	-0.110***
	(-5.57)	(-5.50)	(-5.53)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	-0.050	-0.013	-0.052
F-test: active-inactive=0	1.526	0.083	0.570
p-value, F-test	0.217	0.773	0.450
Mean dep. var	4.355	4.354	4.355
Pseudo R-squared	0.079	0.079	0.079
No. of observations	771,584	769,627	763,253

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01.*

Table A14: Heterogeneity – Time spans (Children ever born)

	(1) Tobit - Children Ever Born	(2) Tobit - Children Ever Born	(3) Tobit - Children Ever Born
Less than 5 years	-0.152***		
	(-4.79)		
Between 5 and 10 years		-0.146***	
		(-3.29)	
More than 10 years			0.063
			(0.88)
Inactive 25 km	-0.050***	-0.051***	-0.050***
	(-4.18)	(-4.21)	(-4.13)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	-0.102	-0.095	0.113
F-test: active-inactive=0	9.233	4.335	2.456
p-value, F-test	0.002	0.037	0.117
Mean dep. var	2.861	2.859	2.862
Pseudo R-squared	0.301	0.301	0.301
No. of observations	841,037	839,050	832,137

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01.*

Table A15: Heterogeneity – Time spans (Current fertility)

	(1) Probit - Current Fertility	(2) Probit - Current Fertility	(3) Probit - Current Fertility
Less than 5 years	-0.021***		
	(-4.51)		
Between 5 and 10 years		-0.011*	
		(-1.81)	
More than 10 years			0.020*
			(1.75)
Inactive 25 km	-0.009***	-0.009***	-0.009***
	(-3.73)	(-3.80)	(-3.76)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	-0.012	-0.002	0.028
Chi2-test: active-inactive=0	5.764	0.087	6.184
p-value, Chi2-test	0.016	0.768	0.013
Mean dep. var	0.266	0.265	0.266
Pseudo R-squared	0.182	0.182	0.181
No. of observations	841,037	839,050	832,137

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01*.

Table A16: Heterogeneity – Abandoned LSLAs and fertility

	(1) Tobit - Ideal No Children	(2) Tobit - Children Born	(3) Probit - Current Fertility
Abandoned 25 km	0.063	0.0122	0.002
	(1.42)	(0.34)	(0.35)
Active 25 km	-0.149***	-0.130***	-0.012***
	(-6.13)	(-5.65)	(-3.72)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	0.212	0.142	0.015
F-test: active-inactive=0	18.47	11.71	
Chi2-test: active-inactive=0			3.852
p-value, F/Chi2-test	0.000	0.000	0.0497
Mean dep. var	4.351	2.853	0.265
Pseudo R-squared	0.079	0.301	0.181
No. of observations	787,763	857,998	857,998

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). **p < 0.10, **p < 0.05, ***p < 0.01*.

Table A17: Robustness – Most accurate spatial level of land deals

	(1) Tobit - Ideal No Children	(2) Tobit - Children Born	(3) Probit - Current Fertility
Active 25 km	-0.342***	-0.154***	-0.006
	(-7.13)	(-4.28)	(-0.85)
Inactive 25 km	-0.379***	-0.189***	-0.006
	(-6.54)	(-5.14)	(-0.87)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	0.037	0.035	0.000
F-test: active-inactive=0	0.240	0.458	-
Chi2-test: active-inactive=0	-	-	0.002
p-value, F/Chi2-test	0.624	0.499	0.969
Mean dep. var	4.348	2.853	0.265
Pseudo R-squared	0.080	0.301	0.181
No. of observations	793,681	864,159	864,159

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01*.

Table A18: Robustness – Migration

	(1) Tobit - Ideal No Children	(2) Tobit - Children Born	(3) Probit - Current Fertility
Active 25 km	-0.343***	-0.092***	-0.007
	(-8.41)	(-2.84)	(-0.90)
Inactive 25 km	-0.098***	-0.028	-0.003
	(-2.60)	(-1.38)	(-0.59)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	-0.245	-0.065	-0.004
F-test: active-inactive=0	20.66	2.940	-
Chi2-test: active-inactive=0	-	-	0.238
p-value, F/Chi2-test	0.000	0.086	0.626
Mean dep. var	4.387	2.704	0.244
Pseudo R-squared	0.074	0.328	0.182
No. of observations	213,060	228,429	228,429

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01*.

Table A19: Robustness – LSLAs area of 15km

	(1) Tobit - Ideal No Children	(2) Tobit - Children Born	(3) Probit - Current Fertility
Active 15 km	-0.140***	-0.117***	-0.014***
	(-4.45)	(-3.21)	(-2.80)
Inactive 15 km	-0.059**	-0.048***	-0.011***
	(-2.45)	(-3.05)	(-3.51)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	-0.081	-0.069	-0.003
F-test: active-inactive=0	4.373	3.068	-
Chi2-test: active-inactive=0	-	-	0.312
p-value, F/Chi2-test	0.0365	0.080	0.576
Mean dep. var	4.351	2.854	0.265
Pseudo R-squared	0.080	0.301	0.181
No. of observations	784,921	855,080	855,080

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01*.

Table A20: Robustness – LSLAs area of 50km

	(1) Tobit - Ideal No Children	(2) Tobit - Children Born	(3) Probit - Current Fertility
Active 50 km	-0.210***	-0.128***	-0.013***
	(-12.26)	(-9.75)	(-6.25)
Inactive 50 km	-0.110***	-0.038***	-0.006***
	(-6.99)	(-4.44)	(-3.80)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	-0.101	-0.090	-0.007
F-test: active-inactive=0	22.31	36.89	-
Chi2-test: active-inactive=0	-	-	7.821
p-value, F/Chi2-test	0.000	0.000	0.005
Mean dep. var	4.349	2.855	0.265
Pseudo R-squared	0.080	0.301	0.182
No. of observations	777,809	847,742	847,742

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01*.

Table A21: Robustness – Control group within 200 km

	(1) Tobit - Ideal No Children	(2) Tobit - Children Born	(3) Probit - Current Fertility
Active 25 km	-0.103***	-0.095***	-0.009***
	(-4.52)	(-4.38)	(-2.90)
Inactive 25 km	-0.078***	-0.023**	-0.004*
	(-4.03)	(-1.97)	(-1.91)
Socio-Demographics	Yes	Yes	Yes
Difference in differences	-0.024	-0.072	-0.005
F-test: active-inactive=0	0.717	8.851	-
Chi2-test: active-inactive=0	-	-	1.738
p-value, F/Chi2-test	0.397	0.003	0.187
Mean dep. var	4.268	2.773	0.255
Pseudo R-squared	0.070	0.310	0.191
No. of observations	602,365	652,136	652,136

Notes: For columns (2) and (3) coefficients are replaced by average marginal effects. T-statistics are in parentheses. Difference-in-differences tests are presented in bottom rows. All regressions control for year and region fixed effects as well as socio-demographic controls as in column (3) of the main findings (Tables 3, 4 and 5). LSLAs being abandoned are excluded from all regressions. **p < 0.10, **p < 0.05, ***p < 0.01.*

A.3 Data appendix

Datasets

The Land Matrix. International Land Coalition (ILC), Centre de Coopéracion Internationale en Recherche Agronomique pour le Développement (CIRAD), Centre for Development and (CDE), German Institute of Global and Area Studies (GIGA) and Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). Web. [accessed August 2020 (investor data updated November 2020 due to data issues in the original data download)]

The DHS Program. Demographic and Health Surveys. United States Agency for International Development (USAID). Web. [accessed August 2020]

Covid-19 Household Telephone Survey. Liberia. 2020. German Institute of Global and Area Studies (GIGA). [accessed December 2020]

Village Chief Telephone Survey. Liberia. 2020. German Institute of Global and Area Studies (GIGA). [accessed December 2020]

DHS data

The DHS women sample includes all geocoded individual records conducted in SSA in the year 2000 or later. Surveys conducted before 2000 were excluded since the Land Matrix database does not provide information about land deals occurring before 2000. Whenever two datasets for the same DHS wave and country are available, the older one is dropped. This applies to survey conducted in Ghana, Malawi, and Rwanda. Tanzania is the exception as only the older version has GPS data. Additionally, the continuous Senegal survey is reduced to one observation per wave.

To obtain information about husband's ideal number of children, when available, the DHS couple records for the respective surveys were merged with the women sample. Considering the sample size of the variable regarding husband's ideal number of children (see Table A9), 171,468 couple observations are included in the sample.

As stated in subsection 3.2, the wealth score was calculated using variables of the DHS standard questionnaire. Based on Croft et al. (2018)'s classifications, several binary variables were constructed. Thereby, binary variables for the source of drinking water, type of toilet and cooking fuel measure whether the source or type can be classified as being 'improved' or 'not improved'.[37] Furthermore, dichotomous variables were created indicating whether the household owns a radio, television, fridge, bike, motorcycle, or car. Since several waves of DHS surveys are included in the sample set, many indicators for assets and services proposed by Rutstein and Johnson (2004) are missing and therefore cannot be included in the analysis. A benefit of using a PCA is that indicator weights are assigned when using this method. The procedure of the PCA is explained in the report written by Rutstein and Johnson (2004). The score was adjusted to contain non-negative values only and starting at zero.

Land Matrix data

The data on land deals were retrieved from the Land Matrix Global Observatory. As defined by the Land Matrix (2020a), to be included in the open-source database, deals have to fulfil the following criteria:

- contain a transfer of rights to use, control or own land through sale, lease or concession;
- been initiated since the year 2000;
- cover at least an area of 200 hectares;
- imply a potential land conversion from small-scale production, local community land use or ecosystem service provision to commercial land use.

Due to the nature of the Land Matrix dataset, a comprehensive data cleaning was needed. Especially concerning the negotiation and implementation status, inconsistent data needed to be replaced. A special thanks goes to Dr Insa Flachsbarth from the GIGA Institute who has shared her comprehensive data cleaning and construction of a panel dataset for the Land Matrix data. Using this panel dataset, it is possible to determine whether and in which year (between 2000 and 2020) the LSLAs deal was intended, concluded as well as whether and when the LSLAs is operating or even abandoned.

[37] The type of floor was excluded from the PCA since for six surveys included in the sample, 99 percent of the responses regarding the type of floor were missing.

Using Flachsbarth's panel data set, land deals conducted in high income countries (according to World Bank classification of July 2012) have been excluded from the sample set as well as land deals including mining activities. Other default filters of the Land Matrix website – such as the exclusion of oil/gas extraction, forest concessions, pure contract farming or domestic deals (Land Matrix, 2020a) – are not applied.

Since precise location details are essential for the applied identification strategy of this analysis, LSLAs with the most inaccurate category of spatial accuracy – country level – are excluded from the dataset.

Telephone survey (Liberia)

To get a first understanding of how LSLAs and fertility might be related to each other, telephone interviews were carried out in Liberia, West Africa, between September until November 2020. Thereby, seven questions regarding fertility were included in a household survey conducted by a GIGA research team. Questions regarding fertility were only asked if the respondent was female. Moreover, interviews with village chiefs of the study areas selected by the GIGA research team were conducted. Following the survey design of both types of interviews will be explained.

Women questionnaire: In total, the sample includes 144 female respondents of reproductive age (15-49 years old). Following the analysis of a previous data collection done by the GIGA research team, the sample set was disaggregated by whether women are living in proximity to one of the two large-scale oil palm plantations. The GIGA research team used a distance cut-off of twenty km around the plantation borders. Thus, all villages within this radius are defined as 'more affected', while the other villages are considered comparatively 'less affected' by the plantations. In total, 84 women are living nearby one of the two selected oil palm plantations in Liberia and are thus assumed to be 'more affected' by the plantations than women of the control group.

Village chief questionnaire: Out of the GIGA baseline dataset, 73 village chiefs which have provided a telephone number were selected. Out of these, 39 village chiefs are living in 'more affected' villages surrounded by one of the plantations whereas the remaining 34 chiefs are living in 'less affected' villages further away from the plantations. Considering 'more affected' villages, 30 village chiefs could be reached and interviewed about the daily town life as well as about how the local community has been affected by the establishment of the oil palm plantation. Since most questions in the questionnaire are related to the presence of the two selected oil palm plantations, village chiefs from 'less affected' villages are not included in this study.

Affidavit

I hereby declare that I wrote this thesis paper independently, without assistance from external parties, and without use of other resources than those indicated. All information taken from other publications or sources in text or in meaning are duly acknowledged in the text. I give my consent to have this thesis checked by plagiarism software.

Mannheim, April 12, 2021

Bianca Dülken