A Double Machine Learning Approach to Estimate Location and Religion Effects on Household Welfare in Ghana

Abstract

This study investigates geographical location and religion effects on welfare (standard of living) of households in Ghana. I apply a recently developed double machine learning (DML) estimator, and provide identification based on the conditional independence assumption. Using a multidimensional survey data set on living conditions of households in Ghana, I estimate penitential outcomes and average treatment effects of geographical location and religious affiliation effect on welfare. My results show that living in the savanna region of Ghana significantly lowers a household's welfare relative to a living in coastal or forest area. I also find that the average standard of living of a Christian household in Ghana slightly exceeds that of a Muslim household, but significantly higher than households that adhere to the traditional African region. Government investments in adaptation, specifically, in irrigation, and also transportation, market access and education in the savanna region can help mitigate the gap in welfare level between residents of the area and other areas.

Keywords: Welfare, double machine learning, poverty, household, location, religion

1 Introduction

A recently active area of methodological research is the synthesis of machine learning (ML) methods with causal inference methods (Knaus, 2018). New methods for estimating average treatment effects, see e.g. (Farrell et al., 2015; Athey et al., 2018; Chernozhukov et al., 2018) and heterogeneous treatment effects, see e.g. (Wager and Athey, 2018) have been suggested (Knaus, 2018). Applied research that uses the proposed causal ML estimators for finding heterogeneous treatment effects are common[1]. However, similar research that employs causal ML estimators for estimating average treatment effects is lacking or rare, even though these estimators have been shown to have the

[1]Examples include: Davis and Heller (2017); Lechner et al. (2017); Bertrand et al. (2017)

potential to improve causal estimation in observational studies (Chernozhukov et al., 2016; Knaus, 2018). This research analyzes geographical location and religion effects on welfare (or standard of living) disparities of households in Ghana using large observational data set.

Two reasons motivate my analysis. First, recent new economic geography (NEG) literature, which suggests that within-country inequalities in standard of living and poverty are the natural result of the development process (Fujita et al., 1999; Puga, 1999; Fujita and Thisse, 2002; Scott, 2009). In particular, the success of any policy targeted at improving shared prosperity across all locations critically depends on what factors drive the observed spatial disparity (Nguyen and Dizon, 2017). Early evidence suggests that market access issues and agglomeration economies are two factors that affect national-level welfare inequalities (Nguyen and Dizon, 2017). In theory, an area that has relatively more natural endowments such as favorable agro-ecological/climatic conditions for agricultural activities, or even just more suitable in its location can be expected to provide an advantage for residence. Empirical research that examines this claim to explain welfare disparities within a single country is lacking so far. Thus, this study fills this gap in the literature.

Second, religion has been identified as a key non-consumption factor that can affect an individual or a household's welfare level since religious believes can affect perceptions about wealth (materialistic versus non-materialistic), work attitudes and economic decisions (Sedmak, 2019). Weber (2012) provided one of the few analysis that connects religion to economic behavior. He attributes the modern advent of capitalism to the Protestant reformation [2]. Other scholars have hypothesized that religion affects economic outcomes through religious doctrines that promote thrift, work ethic, honesty and trust McCleary et al. (2003). Stark and Finke (2000) and McCleary et al. (2003) are examples of empirical research that examine the causal influence of religion on individual behaviour and on the determinants of growth at the national, respectively. Ghana, much like the rest of the sub-Saharan African region is highly religious. Faith plays a central role in life's of Ghanaians and has even affected settlement patterns in the country: Majority Christian in the South and Majority Muslim in the North. I investigate the proximate causal effect of religious affiliation on welfare disparities in Ghana.

[2]Adam Smith in "The Wealth of Nations" laid the foundations to make connections between economics and religion stating that religious institutions like any other sector of the economy, are subject to market forces, incentive and competition problems (Smith, 1937).

The conditional independence assumption (CIA) approach is the workhorse for causal effect estimation in empirical economics research. In addition to stating a specific functional form in an estimation, this approach typically requires introducing a large set of control variables in an estimation to obtain plausible results, especially when estimating causal parameters such at average treatment effects in observational studies (Knaus, 2018). In instances where the set of controls variable are allowed to include polynomials and interactions of the base variables, the number of potential controls can easily exceed the number of observations thus, creating a high-dimensionality problem. Routinely used econometric methods for dealing with high-dimensional issues are quite rigid, which can make model selection difficult (Varian, 2014; Knaus, 2018). Chernozhukov et al. (2016) shows that estimating causal effects can be broken down into several prediction problems. This approach is known as double/biased machine learning (DML). The method employs techniques developed in machine learning literature for solving high-dimensional prediction problems (Hastie et al., 2009), and combines that with causal effect estimation methods to carry out estimation on observational data in a way that controls for selection bias in an objective and data-driven fashion. I employ the DML estimation approach in this study.

This research contributes to three strands of literature. First, I add to the economic geography literature and a relatively small section of research on economics of religion. I analyze a unique household level survey data set published by Ghana Statistical Service, which contains variables that make identification and study of causal effects of both location and religion on welfare feasible. Second, I contribute to the nascent causal ML literature, see example, (Athey and Imbens, 2015; Chernozhukov et al., 2016; Linden and Yarnold, 2016; Athey and Imbens, 2017). Despite triggering several methodological contributions, example, (Mackey et al., 2017; Athey and Wager, 2017; Luo and Spindler, 2017; Antonelli et al., 2019)), applications of the DML method in economics research are rare [3]. Thus, I add to the short list of applied empirical studies the use the approach.

My results show that households in the Savanna region of Ghana potentially have the lowest average welfare followed by those in the forest and then coastal regions. Residents of Accra, the capital of Ghana, have the greatest potential average welfare level, even though the disparity (stan-

[3] Knaus (2018) is one of the few known applied studies that employs the DLM estimation approach. The study estimates effects of musical practice on student's skills using an observational data set from the German National Economic Panel Study (NEPS).

dard deviation) is also greatest. The Savanna areas of Ghana have the least favorable agro-climate conditions, and given that that nearly 50 percent of Ghana's population is depend on the sale of farm produce for income, residents of the savanna regions are at a disadvantage relative to the residents of the coastal and forest regions. I also find potential average welfare to be higher for Christian households than Muslim households, non-religious households and households that adhere to the traditional African religion. This later finding at first appears to fall in line with the thesis of Weber (1958) that the Protestant Ethic (Majority, nearly 70 percent of Christians in Ghana are Protestant) serves as a launchpad for economic success even in a developing country like Ghana. Yet geography and education also provide an explanation; approximately half (48 percent) of of Muslims households live in the Savanna area of northern Ghana, which happens to be the least endowed and less developed part of the country. Most adherents of the traditional African religion in Ghana have no formal education and also live in rural areas of the country. Thus, not surprisingly, my results show that households that adhere to the traditional African religion are much worse off in terms of potential average welfare level relative to households in other religious groups.

The rest of the paper proceeds as follows. In the next section I present a brief summary of previous research on model selection for causal estimation and build up to the DML literature. Section 3 provides a brief background on Ghana to create context. Section 4 describes the data. Section 5 discusses the DML approach and Identification. Section 6 presents the estimation procedure and empirical results. Section 7 discusses the results. I conclude in Section 8.

2 Review of literature

Finding estimators using the conditional independence assumption often require stating either a conditional expectation model, a conditional treatment probability (propensity score) model, or both (Knaus, 2018). Not much direction is available on how to choose the "best" model[4]. Model selection gets even more complicated in estimations that potentially face high-demisionality issues,

[4] Hansen (2005) identifies four conceptual issues underlying the standard econometric model selection methods: parametric vision, the assumption that the data generating process is true, carrying out evaluations based on fit, and not accounting for the impact of model uncertainty on inference. He notes that econometric model selection methods should be based on a semiparametric vision, and that models should be taken as approximations, evaluated based on their purpose, and model uncertainty incorporated into inference methods.

especially where the set of control variables include polynomial terms and interactions of the base variables (which is common in observational studies). High-dimentionality issues make obtaining propensity scores required in unconfounded treatment assignments problematic (Knaus, 2018). One model selection approach proposed by Hirano and Imbens (2001) involves retaining only variables that are statistically significant at a pre-set propensity score level. However, Knaus (2018) notes that this method is impractical where the set of possible controls is large (high - dimensional in nature). Rosenbaum and Rubin (1984) and Dehejia and Wahba (1999, 2002) propose a different model selection approach that involves interactively introducing variables into a propensity score model until a satisfactorily balanced treatment and control groups of co-variate distributions are achieved. Newer methods involve using machine learning techniques to identify propensity scores in a estimation (Lee et al., 2010; Wyss et al., 2014). In all these methods, model selection is done post propensity score determination. Meaning, the methods effectively avoids the model selection step in the inferential analysis of observational data (Knaus, 2018).

Knaus (2018) notes that eluding the model selection step in an inferential analysis can be problematic for two reasons. First, "post-model-selection estimators" may lead to invalid statistical inference (Leeb and Pötscher (2005, 2008)). Leeb and Potscher show that inference procedures that come after model selection are not uniformly consistent, which is required for asymptotic properties of estimators to work as approximations in finite samples Knaus (2018). Thus, inference procedures that circumvent the model selection step in finite samples can be incorrect. The second issue arises when only one model (which can be either the outcome model or propensity score model) is used in the model selection step. Belloni et al. (2014a,b) illustrates that such "single-equation approaches" can deliver to misleading statistical inference results because of the simple fact that the conditional independence assumption necessitates controlling for variables that affect both the treatment probability and the outcome variable. Single-equation approaches fail to capture this.

To overcome these problems, Belloni et al. (2014b) and Farrell et al. (2015) propose a method that extends on Hahn (1998) to frame the conditional expectation of an outcome variable and propensity scores as high-dimensional nuisance parameters. The idea then is to obtain high-quality approximations of the nuisance parameters using machine learning techniques (Knaus, 2018). Combining efficient score and high-quality prediction methods delivers uniformly valid inference after model selection (Belloni et al., 2014b; Farrell et al., 2015). One achieves this because rather than pursuing precise variable selection in a single equation model of either the outcome variable or the propensity score model, model selection is done via a high-quality approximation of all nuisance parameters.

Belloni et al. (2017) extends this idea to include all parameters that are identified from moment conditions that satisfy Neyman orthogonality (Neyman, 1959) cited in (Knaus, 2018). Nuisance parameter approximations that satisfy such moment conditions are free of small sample errors. Chernozhukov et al. (2016) named this estimation method the DML method and proceeds to provide various machine learning algorithms can be used to undertake causal inference in this framework. My study applies the DML method.

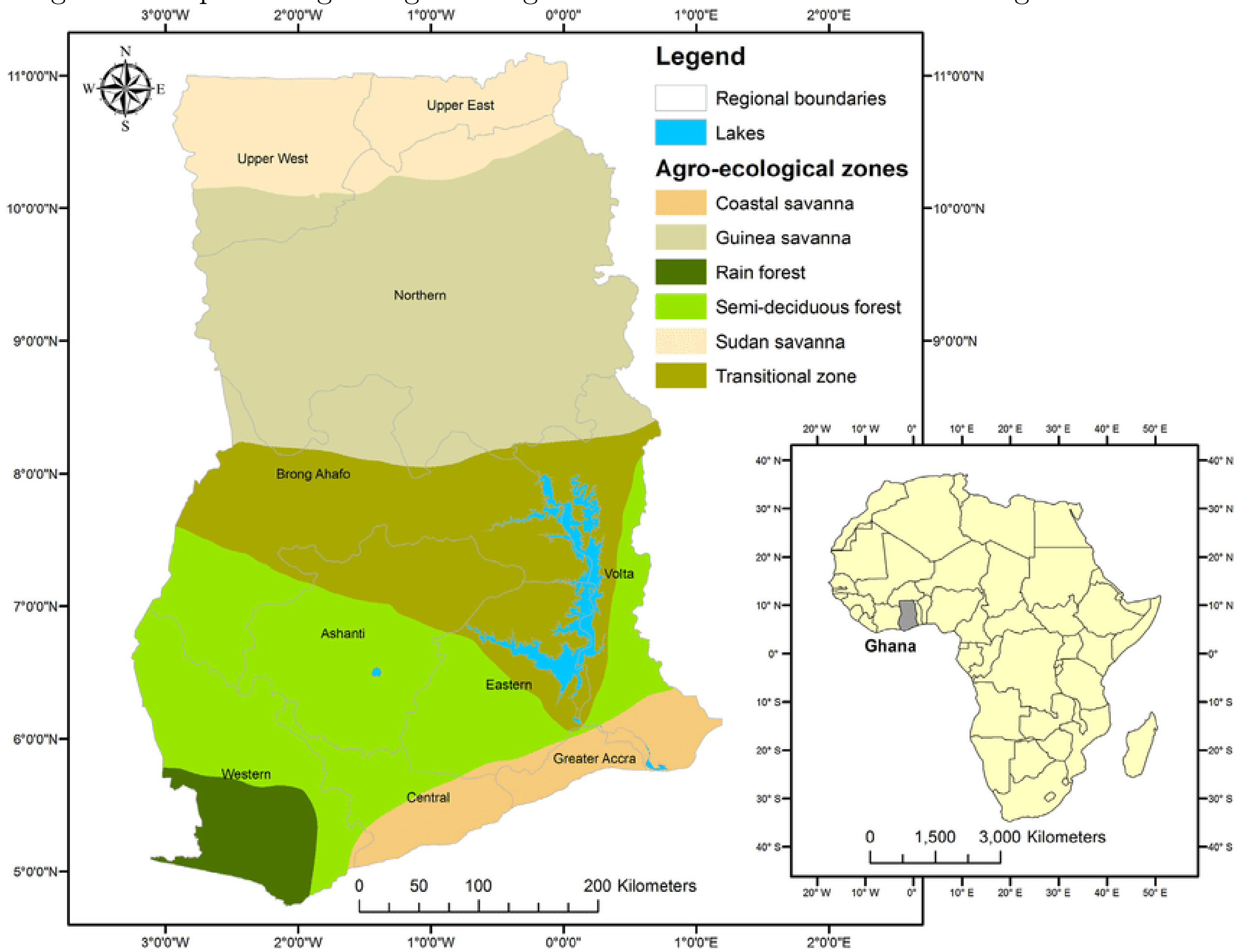
3 Background on Ghana

Here, I provide a brief background on Ghana, highlighting the main administrative regions, agro–ecological zones, and religious groups.

Ghana is located in the west coast of Africa, bordered in the north by Burkina Faso, in the east by Togo, in the west by Ivory Coast and south by the Gulf of Guinea. Ghana's total land area is approximately 148, 197 square miles (or roughly the size of the State of Oregon), and has a population of about 30 million in 2018, giving an overall population density of about 202 persons per square mile. The country lies between latitudes 110 11'N, and 40 44'N and between longitudes 1° 12'E and 30 15'W, making it close both to the Equator and to the Greenwich Meridian. Its proximity to the Equator means relatively high temperatures (78.08°F - 84.2°F) felt in all parts of the country throughout the year. Annual rainfall ranges from about 39 inches in the north to 79 inches in the south. Ghana is divided into ten administrative regions, and may also be classified into four key climatic regions which, define its vegetation: the southwestern equitorial rainforest zone; the west and middle semi-equitorial forest zone; the coastal savannah grassland; and the hot savannah woodland of the northern part of the country (See map in Figure 1)

Geographical location and agro-climatic conditions influence economic activities in Ghana. The forest region, the southwestern equatorial rain forest and the coastal savannah grassland area are all suitable for growing cocoa and many other tropical crops. The coastal area has a relatively

Figure 1: Map showing six agro-ecological zones and ten administrative regions in Ghana

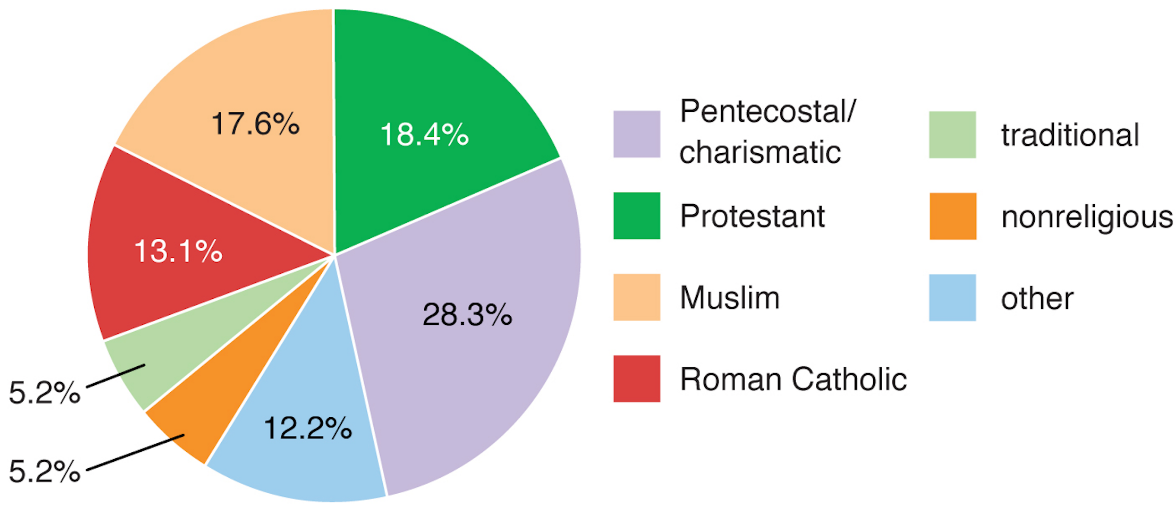


Source: https://www.researchgate.net/figure/Fig-A1-Map-showing-six-agro-ecological-zones-in-Ghanafig6307569254

developed transport system, and prevalence in service, trade, fishing and manufacturing activities. The North, which is predominantly made up of the Guinea savanna and Sudan Savanna vegetation has less developed transportation infrastructure and commerce, and depends heavily on subsistence agriculture. In contrast, the forest and rain forest areas have a diversified structure of production, and traditional agriculture is fast adopting improved technologies.

In terms religion, over 50 percent of Ghana's population is Christian, close to 20 percent is Muslim, and approximately 10 percent either adheres to the traditional African religion or are nonreligious (see Figure 2). The traditional African religion is widespread and deep-rooted, despite not having a systematic set of doctrines. Adherents of this religion worship lesser deities, but also belief in the existence of a supreme being. Significant relevance is given to dead ancestors, who are seen as ever-present and capable of altering the course of events for the living and also capable of serving as the link between the living and the lesser gods.

Figure 2: Religious affiliation in Ghana (2010 National Census)



Source: https://www.britannica.com/place/Ghana/Economy

4 Data

The data used in this study comes from the seventh (latest) round of the Ghana Living Standards Survey (GLSS7) – a multidimensional national household survey conducted by the Ghana Statistical Service in collaboration with the World Bank. The survey collects information on many different aspects of living conditions, including education, health, employment, and household expenditure on food and non-food items. Seven rounds of the survey (in three-year intervals) have been conducted since 1987/88, with the most recent one in 2016/17. The Ghana Statistical Service has consistently published STATA and SPSS files of all data obtained since the first round of the GLSS, freely available for public use, hence the source of my data. Table 6 in the Appendix reports the list of all variables and their descriptions used in this study. A total of 13802 observations across 32 variables is analyzed.

The outcome variable – welfare is the mean value (expressed in US $) of the standard of living measure: total household consumption expenditure per adult equivalent, in the constant prices of

Accra (the Capital of Ghana) in January 2017 [5].

Information on household location in the data is measured in three variables: The first classifies households according to urban - rural placements (i.e 1 = urban, 2 = rural); the second classifies households according to the agro-ecological features in the area (i.e. 1 = coastal, 2 = forest, and 3 = savanna); and the third classification considers both urban - rural placement and the agro-ecological features in the area (i.e. 1 = urban coastal, 2 = urban forest, 3 = urban savanna, 4 = rural coastal, 5 = rural forest, 6 = rural savanna). I present analysis for all three location definitions in this study.

The data on religious affiliation is categorical with the following classification: 1 = No religion, 2 = Catholic, 3 = Protestant / Pentecostals, 4 = Charismatic, 5 = Other christian, 6 = Islam, 7 = Traditionalist, 8 = Other. Table 1 reports summary statistics for numerical variables in the dataset. While Figures 2 and 3 respectively (in Appendix) show density plots of household welfare distributions across different locations.

Table 1: Descriptive statistics

Variable	Mean	Std	Min	Max
Household size	4.20	2.86	1	28
Age of head	46.27	15.90	15	99
Number of dwelling rooms	2.20	1.66	1	33
Number of sleeping rooms	1.84	1.23	1	16
Welfare (in USD)	1046.63	1149.57	9.08	46428.61

[5] Total household consumption expenditure covers food and non-food items, including housing. The standard of living measure accounts for differences in geographical areas, cost of living across all ten administrative regions of Ghana, and household size. The adult equivalence is derived by dividing the total household consumption with the number of adult equivalents in the household.

5 Methodology: the DML approach

5.1 Partially Linear Model

The estimation problem considered by the DLM approach follows a partially linear formulation taken from Robinson (1988). The partially linear specification is

$$Y = \theta_0 D + g_0(X) + U, \quad \mathbb{E}[U|X,\ D] = 0 \tag{1}$$

$$D = m_0(X) + V, \quad \mathbb{E}[V|X] = 0 \tag{2}$$

where Y is the outcome variable (welfare); D is the treatment or target variable (location); X is a high-dimensional vector of covariates or controls; U and V are disturbance terms. θ_0 is the parameter of interest (the average treatment effect), g_0 and m_0 are nuisance parameters [6]. Where $m_0 \neq 0$, typically the case in observational studies, but tends to disappear(or approach zero) in randomized controlled studies (Chernozhukov et al., 2016).

5.2 A Naïve Estimator

A Naïve way to estimate θ_0 in Equation 1 is via a predictive based estimation approach where D and X used to get $\hat{\theta}_0 D + \hat{g}_0 X$. This is achieved by first making an initial guess about $\hat{\theta}_0$, running a Random Forest [7] estimation of $Y - \hat{\theta}_0$ on X to fit $\hat{g}_0 X$, and then OLS of $Y - \hat{g}_0 X$ on D to fit $\hat{\theta}_0$, and repeat until convergence. The estimator for θ_0 in this case is

$$\hat{\theta}_0 = \left(\frac{1}{n} \sum_{i \in I} D_i^2 \right)^{-1} \frac{1}{n} \sum_{i \in I} D_i \left(Y - \hat{g}_0 X \right), \tag{3}$$

which is shown to have excellent prediction performance, but constitutes a heavily biased estimator of θ_0 (i.e. $\hat{\theta}_0 - \theta \neq 0$) (Chernozhukov et al., 2016). Therefore any inferential interpretation of $\hat{\theta}_0$ can be misleading. Chernozhukov et al. attributes the bias in $\hat{\theta}_0$ regularization. Meaning, the g_0 part of the estimation is heavily regularized down towards zero in order to optimize prediction during

[6] Estimation of the nuisance parameters in this case require using machine learning techniques due to the nonparametric nature of the variables in X (Ketzler and Morishige, 2019).

[7] The approach is great for linear approximations of functions that are a combination of trees. So given that that X is generated at a linear combination of trees, the Random Forest method works well

estimation; which leaves a non-trivial effect on the estimation of θ_0. The DML procedure outlined below therefore effectively eliminates regularization bias in the naive estimator.

5.3 DLM Algorithm

The DLM approach encompasses three key steps:

1. Predict Y and D using X by estimating $\mathbb{E}[\hat{Y}|X]$ and $\mathbb{E}[\hat{D}|X]$, using Random Forest or any other good performing Machine Learning Methods.
2. Obtain the residuals: $\hat{W} = Y - \mathbb{E}[\hat{Y}|X]$ and $\hat{V} = D - \mathbb{E}[\hat{D}|X]$.
3. Regress $\hat{W}$ on $\hat{V}$ to get θ^*

Step 1 involves carrying out two prediction exercises; predict Y and D independently using Z, hence the name "Double machine learning". Steps 2 and 3 are fairly straight forward [8]. The new estimator θ^* is unbiased, but is shown to have a larger variance than the biased naïve estimator $\hat{\theta}$. The risk of this former estimator however, is lower hence constitutes a better estimator (Chernozhukov et al., 2016). Moreover, under mile conditions, the estimator, θ^* is $\sqrt{n}$ consistent and approximately centered normal quite generally. A key ingredient in implementing the DML method is sample splitting. Sample splitting helps achieved full efficiency from data use (see Chernozhukov et al. (2016) for details).

6 Empirical Estimation

Before presenting my results, the It is critical that I first address issues of exe in my target variables (location and region). A key issue to address in estimating location effects is endogeneity. If observed agro-ecological features in a particular location is the result of a set of non-random actions such as human activities, then the agro-ecological characteristics of a location cannot be taken as exogenous, and any estimated effects will be biased. The agro-ecological features of different locations in Ghana are completely exclusively the result of nature, and human activities such as deforestation have little

[8] The estimation approach outlined in steps 1-3 is a Frisch-Waugh-Lovell (1930) - style of estimation approach for target parameters.

to no can effect no these features. Specifically, the soil, vegetation cover, and climate of each location is entirely a random assignment of nature. Thus, I take agro-ecological characteristic assignments as exogeneous.

In my estimation, household welfare is my outcome variable, Y, and location defined by the type of agro-ecological features present is my treatment variable, D. I perform two separate estimations for the primary and secondary treatment variables described in section (4). The vector of covariates, X, are provided in Table 1.

I report estimates of the average treatment effect (ATE) of location on welfare. Results on pairwise differences in welfare for all locations are also reported. All estimations are base on sample-splitting as described in section (5.4) applying a 50-50 split. I also report results based on three different methods for estimating the nuisance functions $g(X)$ and $m(X)$ used in forming the othorgonal estimation equations. I consider two l_1- penalization based methods, labled "Lasso" and "Post Lasso", and one tree-based, labled "Random Forest". For the l_1-penalization based method, I use a set of 325 potential control variables formed from the raw set of covariates and all second order terms, i.e. squares and first-order interactions. The results using "Random Forest" are obtained by estimating each nuisance function with a random forest using default settings as in the DML package in R.

Turning to my results, first, estimates under the columns headed "Random Forest" are meant to provide robustness checks for the "post Lasso" estimates, my main results. The potential outcome estimates in Table 2a shows that welfare of a household located in an urban area potentially exceeds per capita welfare of a rural household by about 10% ((1115.4/1022.1 x 100). This difference in welfare is highly statistically significant as reported in Table 2b. The average effect of living in a rural area relative to an urban area is a \$ 93.2 decrease in welfare, give or take \$ 27.9. One explanation for the gap in urban - rural standards of living is that, unlike rural dwellers, urban households are more likely to have a access to electricity, paid water supply, more expensive schools and so on, thus, are more likely to spend more.

In terms of location (classified by agro-ecological features, and Accra) effects on welfare, first, the estimates in Table 3a shows that a household that lives in Accra, the capital, would likely have a welfare level that is approximately 29 %, 30 % and 70 % higher than households in the coastal,

Table 2a: Estimated potential outcomes of location (classified by urban - rural) effects on household welfare (in US $ per year)

Location	Post Lasso	Random Forest
Urban	1115.4	1193.3
	(19.6)	(16.3)
Rural	1022.1	1069.6
	(21.9)	(17.6)

Notes: Estimated potential outcomes reveal welfare (or standard of living) outcomes that would likely be observed if a household was located in a rural or urban area. Robust standard errors (in parenthesis). Column labels denote the method used to estimate nuisance functions.

Table 2b: Estimated difference in average effects of rural versus urban households on welfare(in US $ per year)

Comparison	Post Lasso	Random Forest
Rural vs urban	-93.2***	-123.8***
	(27.9)	(22.1)

Notes: Robust standard errors (in parenthesis). *, ** and *** denote significance at 10 % -, 5 % -, and 1 %, respectively. Column labels denote the method used to estimate nuisance functions.

forest and savanna areas outside Accra respectively. In other words, living in a coastal, forest or savanna area outside Accra posses a statistically significant average effect on welfare, i.e. annual household spending (or welfare) is respectively, \$ 456.3, \$ 483.6 and \$ 867.3 less than welfare in Accra, as shown in Table 3b. The biggest disparity in welfare across agro-ecological areas is between coastal and savanna. Average household spending is \$ 402.0 less in savanna than the coastal region. One explanation for this difference is that most coastal dwellers in Ghana are employed in the non-agriculture sector which pays more. The good climate, proximity to the sea and the capital attracts factories and businesses to locate in the coast. Unlike people in coast, most households in the savanna region especially up north are employed in subsistence agriculture. Low rainfall levels and less favorable climate in the savanna region makes agriculture less profitable, despite most of the agricultural land.

Table 3a: Estimated potential outcomes of location (classified by agro-ecological features, and Accra) effects on household welfare (in US \$ per year)

Location	Post Lasso	Random Forest
Coastal	1630.0	1685.9
	(39.7)	(42.2)
Forest	1611.6	1648.0
	(34.5)	(37.0)
Savanna	1228.0	1585.4
	(119.2)	(85.9)
Accra	2095.3	2161.1
	(137.9)	(104.2)

Notes: Estimated potential outcomes reveal welfare (or standard of living) outcomes that would likely be observed if a household was located in a coastal, forest or savanna area, or Accra. Robust standard errors (in parenthesis). Column labels denote the method used to estimate nuisance functions.

The results in Tables 4a and 4b consider location effects that explicitly account for the urban - rural placements of households in addition to the agro-ecological features of the area. Here, the estimated potential outcomes in Table 4a shows that a household located in rural savanna would

Table 3b: Estimated differences in average effects of location on household welfare

Comparison	Post Lasso	Random Forest
Coastal vs Accra	-456.3***	-472.7***
	(142.3)	(128.1)
Forest vs Accra	-483.6***	-522.6***
	(140.7)	(128.0)
Savanna vs Accra	-867.3***	-898.7***
	(181.5)	(163.6)
Forest vs Coastal	-18.3	-50.1
	(50.3)	(38.6)
Savanna vs Coastal	-402.0***	-406.173**
	(125.1)	(117.8)
Savanna vs Forest	-383.7***	-396.119**
	(123.5)	(117.2)

Notes: Robust standard errors (in parenthesis). *, ** and *** denote significance at 10 % -, 5 % -, and 1 %, respectively. Column labels denote the method used to estimate nuisance functions. Estimates are in US $ per year.

likely record the lowest standard of living ($ 1, 001.2 per year) with a standard deviation of ($ 60.5), followed for by a household in urban savanna ($ 1, 175.2 per year) with a standard deviation of ($ 72.9), rural forest ($ 1, 406.7 per year) with a standard deviation of ($ 56.9), rural coastal ($ 1, 475.9) with a standard deviation of ($ 75.3), urban coastal ($ 1539.2 per year) with a standard deviation of ($ 42.3), urban forest ($ 1609.7 per year) with a standard deviation of ($ 41.2) and Accra ($ 2123.6 per year) with a standard deviation of ($ 145.7) accordingly. Notice that Accra has the biggest spread in welfare distribution. Table 4b, shows test results for pairwise differences in average effects between locations. The largest statistically significant average effect disparity in welfare is between rural savanna and Accra, estimated at about ($ 862.3 per year), give or take $ 157.6 per year. The second biggest difference average location effect is between rural savanna and urban forest, ($ 851.6), with a standard deviation of ($ 217). Meaning, a household located in rural savanna spends on average $ 851.6 less than a counterparts in urban forest region. The same reasons given earlier possibly account for this disparities.

Moving on, Table 5a shows potential outcome estimates for religious affiliation effects on welfare. Estimated potential average welfare is smallest ($ 624.4 per year) with a standard deviation of ($ 36.7) for households affiliated with the traditional African religion (Traditionalists), followed by households affiliate with other religions ($ 700 per year), non religious households ($ 732.0 per year), Islamic households ($ 784.9 per year), other christian ($ 801 per year), Catholic ($ 805.5 per year), Protestant or Pentecostal ($ 819.1 per year) and Charismatic ($ 854.7 per year). The biggest statistically significant disparity in average welfare is between traditionalist and Protestant/Pentecostal households ($ 229 per year) from Table 5b. In general, pairwise comparisons of difference in average welfare among households affiliated with the Catholic, Protestant/Pentecostal, Charismatic, Islamic and other Christian faiths did not reveal any statistically significant differences, see results in Table 5b. However, the difference in average welfare between households affiliated with the Traditional African religion (Traditionalist) and all other faith categories considered was significant. Nearly 5.2% of Ghana's populace identify as traditionalists according to Ghana's 2010 population a and housing census. Most of these households about (73 %) according to the census are uneducated and live in rural areas. Which would seem to explain why households affiliated with this faith have a significantly lower welfare on average.

7 Discussion

Other things being equal, consumption of more goods and services increases a person's welfare level. This claim is largely supported by much of observed human behavior. Certainly, there are many other factors that affect welfare besides consumption of goods and services, but since these other factors tend to be much more difficult to measure, economists typically limit themselves to that part of human welfare that directly deals with consumption. Restricting welfare measurement to consumption only, means that a considerable number of non-consumption factors are left out. This then affects the kinds of policies and their effectiveness designed to improve welfare (standard of living) and reduce inequality. I study the effects of two non-consumption factors that affect welfare: location, see (Okwi et al., 2007; Mishra, 2011; Annim et al., 2012; Nguyen and Dizon, 2017) and religion, see (Beyers, 2014; Sedmak, 2019), using a DLM approach recently developed for estimating causal effects from observational data.

Table 4a: Estimated potential outcomes of location (classified by rural/urban interacted with agro-ecological features of area, and Accra) effects on household welfare

Ecological zone	Post Lasso	Random Forest
Urban coastal	1539.2	1636.4
	(42.3)	(51.0)
Urban forest	1609.7	1575.0
	(41.2)	(47.3)
Urban savanna	1175.2	1167.1
	(72.9)	(108.7)
Rural coastal	1475.9	1371.2
	(75.3)	(61.7)
Rural forest	1406.7	1484.7
	(56.9)	(62.7)
Rural savanna	1001.2	1097.7
	(60.5)	(59.4)
Accra	2123.6	2165.2
	(145.7)	(110.4)

Notes: Estimated potential outcomes reveal welfare (or standard of living) outcomes that would likely be observed if a household was located in urban/rural coastal, forest or savanna areas, or Accra. Estimates are in US $ per year. Robust standard errors (in parenthesis). Column labels denote the method used to estimate nuisance functions.

Table 4b: Estimated differences in average effects of location on household welfare

Comparison	Post Lasso	Random Forest
Urban coastal vs Accra	-584.3***	-598.9***
	(150.2)	(139.9)
Urban forest vs Accra	-513.9***	-540.2***
	(149.9)	(142.5)
Urban savanna vs Accra	-748.4***	-750.1***
	(162.2)	(153.8)
Rural coastal vs Accra	-647.6***	-664.1***
	(162.5)	(150.1)
Rural forest vs Accra	-716.9***	-730.6***
	(155.2)	(142.4)
Rural savanna vs Accra	-862.3***	-883.5**
	(157.6)	(149.7)
Urban forest vs Urban coastal	70.4	81.4
	(56.7)	(46.5)
Urban savanna vs Urban coastal	-364.1***	-371.3**
	(83.5)	(69.7)
Rural coastal vs Urban coastal	-63.341	-58.7
	(84.436)	(78.3)
Rural forest vs Urban coastal	-132.6*	-141.7*
	(69.3)	(78.5)
Rural savanna vs Urban coastal	-522.3	-527.7
	(217.0)	(185.5)
Urban savanna vs Urban forest	-434.5***	-457.9**
	(82.850)	(67.1)
Rural coastal vs Urban forest	-233.8	-243.8
	(84.3)	(75.7)
Rural forest vs Urban forest	-203.0***	-219.324***
	(68.7)	(75.6)
Rural savanna vs Urban forest	-851.6	-866.7***
	(217.0)	(184.2)
Rural coastal vs Urban savanna	300.7***	315.9***
	(104.1)	(95.9)
Rural forest vs Urban savanna	231.5**	238.6
	(91.9)	(88.4)
Rural Savanna vs Urban savanna	-286.1**	-319.6*
	(225.3)	(209.1)
Rural forest vs Rural coastal	-69.3	-50.5
	(93.5)	(86.5)
Rural savanna vs Rural coastal	-68.3	-51.5
	(125.9)	(118.8)
Rural savanna vs Rural forest	-154.6	-166.0
	(122)	(111.1)

Notes: Robust standard errors (in parenthesis). *, ** and *** denote significance at 10 % -, 5 % -, and 1 %, respectively. Column labels denote the method used to estimate nuisance functions. Estimates are in US $ per year.

Table 5a: Estimated potential outcomes of religious affiliation effects on household welfare (in US $ per year)

Religious affiliation	Post Lasso	Random Forest
No religion	732.40	741.2
	(47.6)	(45.1)
Catholic	805.5	825.6
	(29.3)	(28.7)
Protestant	819.1	859.1
	(32.0)	(29.2)
Charismatic	854.7	836.2
	(22.4)	(17.5)
Other Christian	801.0	805.4
	(38.0)	(34.9)
Islam	784.9	787.6
	(48.7)	(44.7)
Traditionalists	624.4	632.3
	(36.7)	(32.2)
Other religion	700.0	696.3
	(103.9)	(361.2)

Notes: Estimated potential outcomes reveal welfare (or standard of living) outcomes that would likely be observed if a household was affiliated with a certain religion. Robust standard errors (in parenthesis). Column labels denote the method used to estimate nuisance functions.

Machine learning (ML) methods developed mostly by computer scientists and statisticians in the last few decades, have achieved extraordinary success in solving prediction problems, especially with high-dimensional, complicated or big data. These methods have also been used with great success, to perform natural language proceeding (e.g. spam filtering) and computer vision, among others.

Prediction problems are however of little interest to economist instead, the problem of measuring causal parameters, including various treatment effects, is much more important. Thus, there has recently been a growing amount of work in the econometrics literature trying to apply ML methods to estimate causal parameters. One such method in the literature is the naive ML estimator which involves naively applying ML methods to estimate causal parameters. This has produced unsatisfactory results, since ML methods, in optimizing prediction become heavily regularized. Thus, ML-based causal parameter estimators become significantly biased, thus delivers sub-optimal precision. Moreover, bias makes it difficult to study distributional properties of estimators and for that matter makes inference overly complicated.

The DML method approach addresses some of the issues of the Naive machine learning estimator. The method is premised on the idea that it is possible to represent a causal parameter of interest as a function of several prediction problems such that the bias in the solution of the prediction problems posses little effect on the causal parameter itself. In fact, so long as such a function can be constructed, the DML method employs ML algorithms to solve each prediction problem separately in a first stage and proceeds to plug the solutions into the function giving the causal parameter of interest in a second stage. Chernozhukov et al show that one can construct the function that links the causal parameter and the solution of the prediction problems by using econometric models via Neyman orthogonal scores, which are have been studied widely in semi-parametric estimation literature. The DML framework offers an opportunity to study economic problems such location and region effects on welfare disparities in a developing country that involve analyzing high-dimensional observational data.

The old African saying that "Tell me where you live, and I can predict how well you'll do in life" points to the fact that location can be an good predictor of a person's welfare Scott (2009). An urban resident in Ghana's capital, Accra, has an 18 percent chance of staying above the national poverty line, and a 98% chance of having access to electricity. By contrast, a rural resident in the northern part of Ghana has more than a 70 percent chance of falling below the national poverty line, and only a 20 percent chance of having access to electricity according to the "Ghana poverty profile report 2005 - 2017". My results show that these disparities in living standards across space in Ghana are not only affected by the urban - rural location of households which is more obvious, but also by the agro-ecological factors of the area which determine the kinds of economic activities and opportunities available. This result is similar to an earlier report by Minot et al. (2006) who finds that nearly 75 % of variance in rural poverty and inequality in Vietnam is explained by variance in agro-climate factors and market access.

Unlike market access, geographic variables such as agro-ecological differences cannot be influenced by policy. However, the need to overcome geographic factors that hinder economic activities often trigger migration. For instance, Van der Geest (2011) reports that one out of every five Northern born Ghanaians live in south. And a survey of migrant farmers from Northern Ghana reveals that the primary reason for moving to the south is environmental. Northern farmers move down south due to poor agro–ecological conditions at home. Minot et al. (2006) argues that to the extent that migration can increase the living standards of migrants without adversely affecting others, migration can be an effective way to tackle welfare and poverty disparities resulting from geographic differences. In reality, migration is not without problems, both in the short and long terms (Asiedu, 2010). The extend to which migration may be more or less problematic depends on the level of training, education and skill set that of the migrant posses. Carefully planned within-country migration policies especially around environmentally induced migrations can help increase the benefits of such migrations for the individual and the nation, more so, minimize standard of living disparities.

Investment in adaptation can be another way that the impacts of unfavorable ago-eocological factors in an area can be minimized on economic activities and living conditions. Unfortunately, this alternative can be very expensive and often not feasible for a developing countries. For instance, the government of Ghana could invest in large scale irrigation as well as subsidize fertilizers to the mostly farm population in the savanna north which which is drier and less fertile. But such a project generally require large amounts of money that the developing economies do not have.

Beyers (2014) discusses the link between poverty and religion. He notes that, religious believes and practice of faith can affect job choice and where a person decides to live. For instance, a devout

Muslim may fail take up a job that involves production and sale of alcohol or pork products, since this is against the Islamic faith. This limits the range of opportunities available to such individuals. I find religious affiliation to have significant average effects on household welfare level in Ghana. Specifically, my results show that welfare to be statistically significantly lowest for households affiliated with the traditional African religion. Most traditionalists in Ghana have no formal education or training and typically live in the rural parts of the country employed as subsistence farmers. Traditionalist in Ghana generally do not migrate because they worship specific objects like tree, lakes or rocks which require staying in a particular location. Thus, seeking greener pastures else is generally not an option for the traditionalist.

8 Conclusion

This study applies the double machine learning method to estimate location and religion effects on welfare (standard of living) of households in Ghana. Location and religion are among the several non-consumption variables identified to affect welfare. Estimating the direct effects of these variables from observational data can present some difficulties including deciding on an appropriate model form and a set of control variables and also dealing with possible issues of high-dimensionality in the estimation. The double machine learning (DML) approach employed in this study is a recent development that is based on a partially linear estimation approach that utilizes machine learning techniques developed for non-linear prediction problems and applies that in an econometric context to obtain estimators with well behaved asymptotic properties. The DML method delivers point estimates that have a $\sqrt{n}$ rate of convergence for N observations and are approximately unbiased and normally distributed.

Using data from the latest (seventh) round of the Ghana living standards survey I find that households that live in the savanna regions of Ghana have a significantly lower standard of living relative to households in the coastal and forest areas. This disparity in living standards is further exacerbated when the urban-rural placements of households is factored into consideration. Investment in adaptation such as irrigation is one way that location impacts on welfare may be curtailed. Also, a carefully designed within-country migration policy that allows migrants to benefit without necessarily creating other problems may help reduce location effects on welfare disparities in Ghana.

I also find religion to have significant effects on welfare disparities in Ghana. Specifically, households affiliated with traditional African religion have average welfare levels significantly lower than Christian and Islamic households. Ghana has been remarkably stable in terms of mutual respect and relations among for the many different religious groups. Except for households affiliated with the traditiional african religion, differences in welfare levels across the different religious groups is marginal. This perhaps suggests that religion may not a decisive factor of the welfare levels of people in Ghana, thus explaining why there have been no problems in the relations between the different religious groups.

Ultimately identifying the effects of non-consumption factors such location and religion on welfare disparities in a developing country like Ghana is crucial to develop holistic strategies to address inequality and increase overall will being of the the people.

An important limitation of this study is the assumption that households posses the same utility function. This assumption is necessary to facilitate comparison across households. In reality households can have different welfare functions and impossible or even meaningless to compare.

References

Annim, S. K., Mariwah, S., and Sebu, J. (2012). Spatial inequality and household poverty in ghana. *Economic Systems*, 36(4):487–505.

Antonelli, J., Parmigiani, G., Dominici, F., et al. (2019). High-dimensional confounding adjustment using continuous spike and slab priors. *Bayesian Analysis*, 14(3):825–848.

Asiedu, A. B. (2010). Some perspectives on the migration of skilled professionals from ghana. *African Studies Review*, 53(1):61–78.

Athey, S. and Imbens, G. W. (2015). Machine learning methods for estimating heterogeneous causal effects. *stat*, 1050(5):1–26.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Athey, S., Imbens, G. W., and Wager, S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):597–623.

Athey, S. and Wager, S. (2017). Efficient policy learning. *arXiv preprint arXiv:1702.02896*.

Belloni, A., Chernozhukov, V., Fernández-Val, I., and Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1):233–298.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Bertrand, M., Crépon, B., Marguerie, A., and Premand, P. (2017). *Contemporaneous and Post-Program Impacts of a Public Works Program: Evidence from Côte d'Ivoire*. World Bank.

Beyers, J. (2014). The effect of religion on poverty. *HTS Theological Studies*, 70(1):01–10.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. K. (2016). Double machine learning for treatment and causal parameters. Technical report, cemmap working paper.

Davis, J. M. and Heller, S. B. (2017). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of Economics and Statistics*, pages 1–47.

Dehejia, R. H. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448):1053–1062.

Dehejia, R. H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1):151–161.

Farrell, S. A., Murphy, T., and Lo, K. K. (2015). Autoclassification of the variable 3xmm sources using the random forest machine learning algorithm. *The Astrophysical Journal*, 813(1):28.

Fujita, M., Krugman, P. R., and Venables, A. (1999). *The spatial economy: Cities, regions, and international trade*. MIT press.

Fujita, M. and Thisse, J.-F. (2002). Agglomeration and market interaction.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, pages 315–331.

Hansen, B. E. (2005). Challenges for econometric model selection. *Econometric Theory*, 21(1):60–68.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hirano, K. and Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes research methodology*, 2(3-4):259–278.

Ketzler, C. and Morishige, G. (2019). Implementation of the double/debiased machine learning approach in python.

Knaus, M. C. (2018). A double machine learning approach to estimate the effects of musical practice on student's skills. *arXiv preprint arXiv:1805.10300*.

Lechner, M., Strittmatter, A., and Knaus, M. (2017). Heterogeneous employment effects of job search programmes: A machine learning approach.

Lee, B. K., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in medicine*, 29(3):337–346.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59.

Leeb, H. and Pötscher, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory*, 24(2):338–376.

Linden, A. and Yarnold, P. R. (2016). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice*, 22(6):875–885.

Luo, Y. and Spindler, M. (2017). L 2-boosting for economic applications. *American Economic Review*, 107(5):270–73.

Mackey, L., Syrgkanis, V., and Zadik, I. (2017). Orthogonal machine learning: Power and limitations. *arXiv preprint arXiv:1711.00342*.

McCleary, R., Barro, R., et al. (2003). Religion and economic growth across countries. Technical report.

Minot, N., Baulch, B., and Epprecht, M. (2006). Poverty and inequality in vietnam: Spatial patterns and geographic determinants. Technical report.

Mishra, A. P. (2011). Geography, poverty and development: an interdisciplinary debate. *Sociedade & Natureza*, 23(2):227–236.

Neyman, J. (1959). Optimal asymptotic tests of composite hypotheses. *Probability and statsitics*, pages 213–234.

Nguyen, N. T. V. and Dizon, F. F. (2017). *The Geography of Welfare in Benin, Burkina Faso, Côte d'Ivoire, and Togo*. World Bank.

Okwi, P. O., Ndeng'e, G., Kristjanson, P., Arunga, M., Notenbaert, A., Omolo, A., Henninger, N., Benson, T., Kariuki, P., and Owuor, J. (2007). Spatial determinants of poverty in rural kenya. *Proceedings of the National Academy of Sciences*, 104(43):16769–16774.

Puga, D. (1999). The rise and fall of regional inequalities. *European economic review*, 43(2):303–334.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, 79(387):516–524.

Scott, A. J. (2009). World development report 2009: reshaping economic geography.

Sedmak, C. (2019). Evidence-based dialogue: the relationship between religion and poverty through the lens of randomized controlled trials. *Palgrave Communications*, 5(1):1–7.

Smith, A. (1937). The wealth of nations [1776].

Stark, R. and Finke, R. (2000). *Acts of faith: Explaining the human side of religion.* Univ of California Press.

Van der Geest, K. (2011). North-south migration in ghana: what role for the environment? *International Migration*, 49:e69–e94.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Weber, M. (1958). The protestant ethic and the spirit of capitalism (new york: Charles scribner's sons, 1958). *Economy and Society: An Outline of Interpretive Sociology. Guenther Roth and Claus Wittich, eds. New York: Bedminster Press.*

Weber, M. (2012). *Max Weber: The theory of social and economic organization.* Martino Publishing.

Wyss, R., Abegg, B., and Luthe, T. (2014). Perceptions of climate change in a tourism governance context. *Tourism Management Perspectives*, 11:69–76.

Appendices

Figure 3: Density plots of welfare distributions by urban-rural placement of households

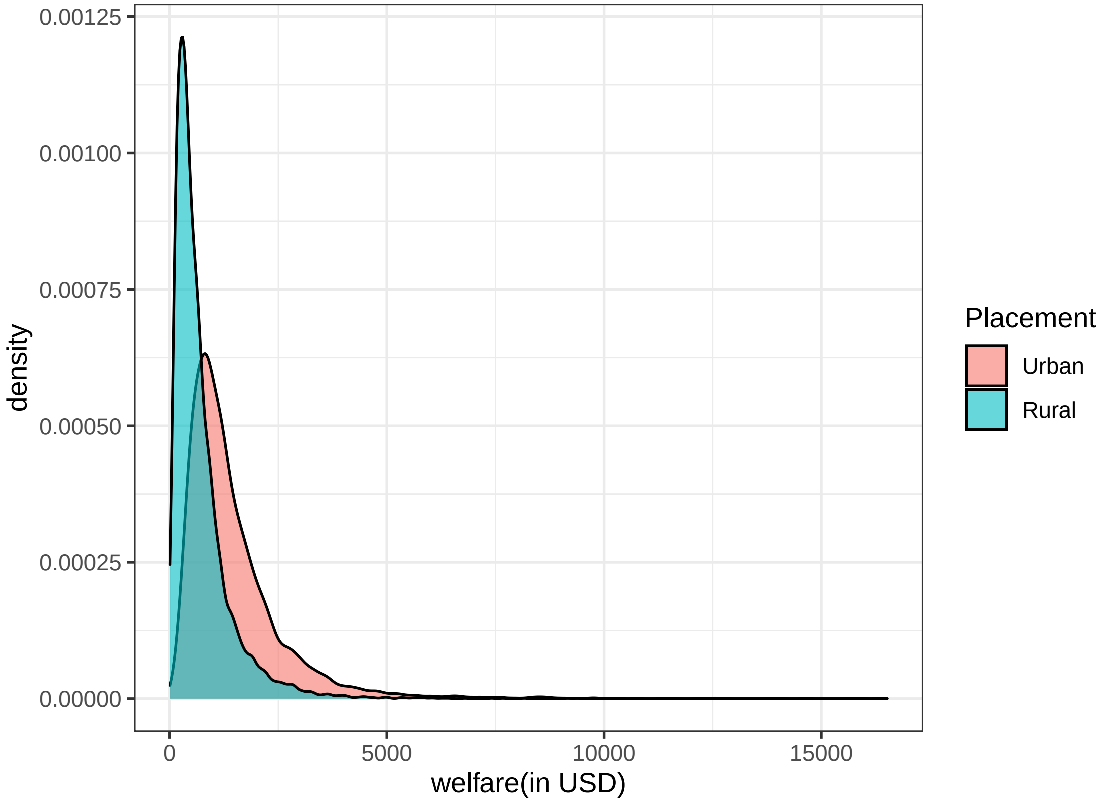


Figure 4: Density plots of welfare distributions across urban-rural/agro-ecological zones and Accra

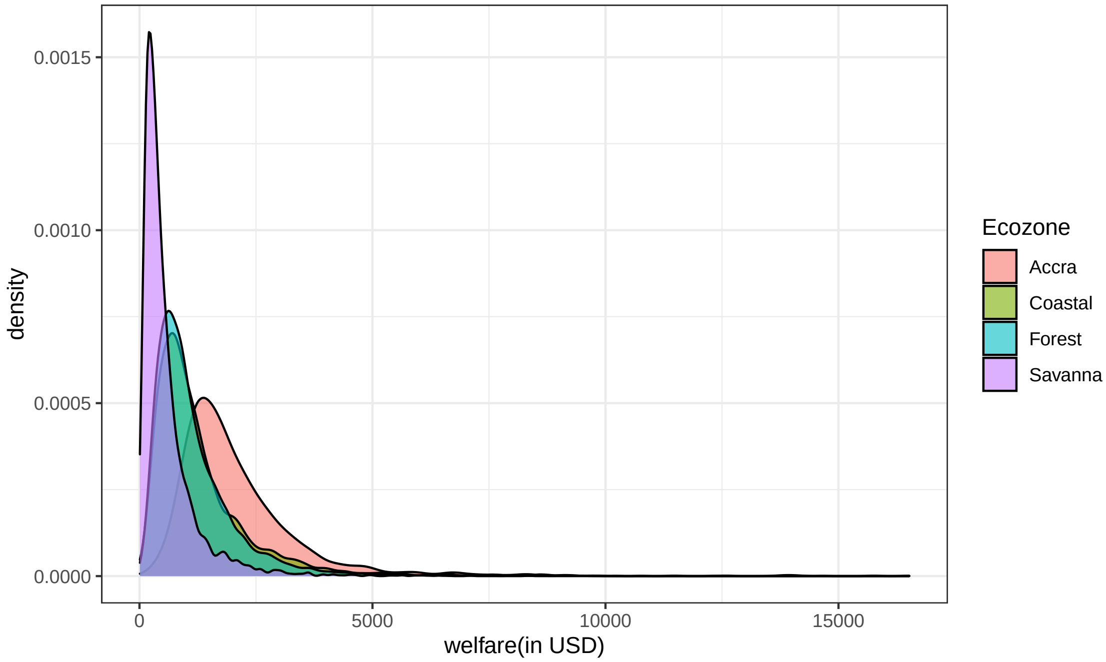


Figure 5: Density plots of welfare distributions across agro-ecological zones and Accra

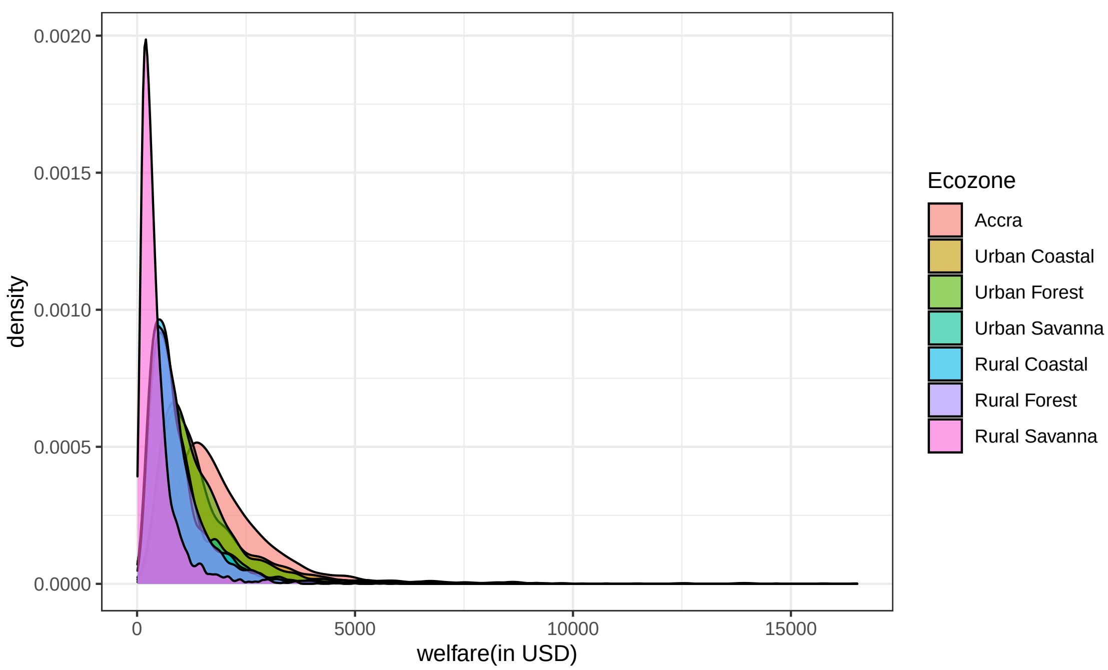


Table 5b: Estimated differences in average effects of religious affiliation on household welfare

Comparison	Post Lasso	Random Forest
Catholic - No religion	82.1	84.4
	(56.2)	(52.6)
Protestant - No religion	100.7**	117.9**
	(54.2)	(52.9)
Charismatic - No religion	90.3***	94.9**
	(52.8)	(47.4)
Other Christian - No religion	60.6	64.1
	(57.2)	(56.3)
Islam - No religion	45.5	46.4
	(64.0)	(62.9)
Traditionalist - No religion	-107.0*	-109.0**
	(71.1)	(55.0)
Other religion - No religion	-47.0	-44.9
	(370.2)	(363.9)
Protestant - Catholic	39.6	33.6
	(46.5)	(39.3)
Charismatic - Catholic	9.2	10.6
	(34.7)	(32.0)
Other Christian - Catholic	-10.5	-20.2
	(46.03)	(44.2)
Islam - Catholic	-39.4	-38.0
	(57.9)	(52.2)
Traditionalist - Catholic	-190.1***	-193.3***
	(56.5)	(42.4)
Other religion - Catholic	-133.1	-129.3
	(353.4)	(362.2)
Charismatic - Protestant	-18.0	-22.9
	(32.0)	(32.3)
Other Christian - Protestant	-55.1	-53.8
	(47.7)	(44.3)
Islam - Protestant	-80.7	-71.5
	(55.9)	(52.4)
Traditionalist - Protestant	-229.8***	-226.9***
	(45.2)	(42.8)
Other religion - Protestant	-168.7	-162.9
	(367.6)	(362.2)
Other Christian - Charismatic	-39.8	-30.8
	(38.1)	(37.9)
Islam - Charismatic	-49.9	-48.5
	(49.5)	(47.0)
Traditionalist - Charismatic	-210.4***	-203.9***
	(36.2)	(35.8)
Other religion - Charismatic	-150.3	-139.9
	(366.0)	(361.5)
Islam - Other Christian	-19.0	-17.7
	(58.8)	(55.9)
Traditionalist - Other Christian	-180.5***	-173.1***
	(47.2)	(46.9)
Other religion - Other Christian	-120.5	-109.1
	(377.8)	(362.8)
Traditionalist - Islam	-160.5***	-155.3***
	(56.2)	(54.6)
Other religion - Islam	-101.5	-91.3
	(370.0)	(363.8)
Other religion - Traditionalist	80.0	64.0
	(371.3)	(362.5)

Notes: Robust standard errors (in parenthesis). *, ** and *** denote significance at 10 % -, 5 % -, and 1 %, respectively. Column labels denote the method used to estimate nuisance functions. Estimates are in US $ per year.

Table 6: Description of variables obtained from the GLSS7

Variable	Type of Variable	Number of Categories	Variable Description
Outcome variable (Y) :			
welfare (in USD)	continuous	-	Welfare (Average standard of living in US dollars)
Variables of interest (D):			
location2-urban-rural	categorical	2	urban or rural location of household: A
ecozone	categorical	3	Agro-ecological classification of location: B
location7-ecozone	categorical	7	Combination of A and B
religious-denom	categorical	8	Religious affiliation of household
Covariates (or nuisance variables) (X):			
hhsize	count	-	Size of household
age-head	continuous	-	Age of economic head of household
gender-head	categorical	2	Gender of economic head of household
marital-status-head	categorical	6	Marital status of economic head of household
ethnicity	categorical	94	Ethnicity of household
region	categorical	10	Region in which household is located
employment-status	categorical	3	Employment status of economic head of household
employment-type	categorical	7	Employment type of economic head of household
highest-education	categorical	6	Highest education of economic head
dwelling-type	categorical	11	Type of dwelling of household
const-mat-outerwall	categorical	10	Construction material of outerwall of dwelling
const-mat-roof	categorical	9	Construction material of roof of dwelling
const-mat-floor	categorical	9	Construction material of floor of dwelling
dwellingrooms-num	Count	-	Number of dwelling rooms in house
sleepingrooms-num	Count	-	Number of sleeping rooms in house
mainsource-of-light	categorical	8	Main source of light for dwelling
mainsource-of-drinkingwater	categorical	16	Main source of drinking water in dwelling
mainsource-of-cookingfuel	categorical	10	Main source of cooking fuel for household
type-of-toilet	categorical	7	Type of toilet in dwelling
own-fixlinephone	categorical	2	If household owned a fix line phone
own-paidphone	categorical	2	If household owned a paid phone

Source: Ghana Living Standards Survey, 2016/17