Master Thesis on

Developing a Bottom-Up Methodology to Project Electricity Demand for Productive Uses in sub-Saharan Africa in 2030 in IMAGE

September 14th, 2020

Student name: Su-Min Choi

Student number: 6619347

Student email: s.i.choi@students.uu.nl

Student address: Willem van Noortstraat 112, 3514 GH Utrecht

Master program: MSc Sustainable Development

Track: Energy and Materials

Supervisor UU: Gert Jan Kramer

Second Reader: Andries Hof

Extent: 30 EC

Word count: 38,283

Internship institution: Netherlands Environmental Assessment Agency PBL (Planungsbureau voor de Leefomgeving)

Internship address: Bezuidenhoutseweg 30, 2594 AV, The Hague

Internship supervisor: Anteneh Dagnachew

Internship supervisor email: Anteneh.Dagnachew@pbl.nl

Utrecht University

PBL Netherlands Environmental Assessment Agency

Table of Contents

List of Figures

List of Tables

Index of Abbreviations

AEI - Africa Electrification Initiative

AEO – Annual Energy Outlook

ANOVA – Analysais of Variance

B – Beta (coefficient)

DHS – Demographic and Health Surveys

DRC – The Democratic Republic of the Congo

EP – Enterprise

EUEI – European Union Energy Initiative for Poverty Eradication and Sustainable Development

ESMAP – Energy Sector Management Assistance Program

GDP – Gross Domestic Product

GHG – Greenhouse Gas

GHS – General Household Survey

GIS – Geographic Information System

GIZ – Deutsche Gesellschaft für Internationale Zusammenarbeit (German Society for International Collaboration)

HDI – Human Development Index

HH – Household

HHH – Household head

IEA – International Energy Agency

IFC – International Monetary Fund

IHDI – Inequality-adjusted Human Development Index

IMAGE – Integrated Model to Assess the Global Environment

kWh – kilo-Watt hours

LN – Logarithmus Naturalis (natural logarithm)

LSMS – Living Standards Measurement Study LSMS

ME – Microenterprise

MSME – Micro-, Small-, and Medium-Enterprises

OnSSET – Open-source Spatial Electrification Tool

PBL – Planbureau voor de Leefomgeving (Netherlands Environmental Asessment Agency)

PPP – Purchasing Power Parity

PV – Photovoltaic

PrElGen – Productive Uses of Electricity Generator model

PRODUSE - Productive Use of Energy

SDG – Sustainable Development Goal

SSA – sub-Saharan Africa

SSP – Shared Socioeconomic Pathway

TIMER – Targets Image Energy Regional Model

UN – United Nations

USD – United Stated of American Dollar

VAT – Value Added Tax

VIF – Variance Inflation Factor

WB – World Bank

WQ – Wealth quintile

“Xamul aay na, laajtewul a ko raw.”

“It is bad to not know, but it is even worse to not ask.”

Lébou Proverb

Abstract

There is a general agreement about the manifold ways in which electricity supply can contribute to sustainable improvement of livelihoods, which is manifested in the SDG 7 of the UN. Many governmental institutions therefore support electrification projects in developing countries with poor access in the context of their development policies. One such institutions is the Dutch Directorate General for International Cooperation, which is aiming to develop a policy strategy to support the electrification of countries in SSA, which currently have the lowest electrification rates in the world.

To support this planning, the IMAGE model, an integrated assessment model by the Dutch Environmental Assessment Agency PBL is supposed to project future electricity demand in the region to understand the consequences of different electrification strategies in terms of costs and technologies. Models like IMAGE are increasingly improving their spatial granularity and accuracy by implementing bottom-up modelling approaches to better understand how underlying premises influence different projection scenarios. To this end, the newly developed PrElGen framework is aiming at reflecting electricity demand from different channels, such as residential demand, demand from schools and health facilities and demand for irrigation and productive uses. The present report explains how a methodology for projecting productive uses was developed, i.e. the use for micro-scale income activities on the HH level of crop processing and non-agricultural productive uses.

The projection for demand from crop processing is based on a literature review of electricity requirements of different processing technologies for a selection of crops carried out in the context of this research. The final model for the projection for other productive uses, which this study developed, consists of four distinct empirical models developed in the course of this research which are two logistic regression models for the propensity of business activity in a household and the propensity to have an electrical connection and two multivariate linear regression for the performance and consumption of enterprises. The regression models were built based on samples from the World Bank Household and Enterprise Surveys and further variables were added from other sources. Despite a multitude of interactions amongst variables, a variety of variables could be distilled for the different models which showed robust predictive power on the dependent variables. The study could thereby contribute to the existing body of research by confirming significant effects over a broad range of context found in previous literature in rather narrow samples. It also adds some new insights into further predictors and analysed interaction effects of several predictors and outlines potential pathways for future research.

Acknowledgements

I want to thank my PBL supervisor Anteneh Dagnachew and Giacomo Falchetta for giving me the opportunity to work on this fascinating project and for their collaboration and support. Further thanks go to my UU supervisor Gert Jan Kramer for his professional and moral support. Of course, I thank my friends and family for their great love, support and patience. Special thanks go to Ines Bouacida, Andréane Jacqueline Barbara Bellon de Chassy and Aurelio Negri for the productive team work on all the professional and non-professional projects. I further want to state my gratitude towards my mother and her colleagues for hosting me so generously in their office facilities as if it was self-evident.

Last but not least, the biggest gratitude goes to my parents for their limitless, unconditional love and support. Words cannot express how blessed I feel to have you.

1 Introduction

1.1 Background

Reliable access to electricity is essential for ensuring decent living standards including modern healthcare, education, telecommunication and general human development as acknowledged by the United Nations (UN) in the form of Sustainable Development Goal 7 (SDG 7): Affordable and clean energy. The different applications for electricity, such as charging of telecommunication devices, spatial cooling and operating machinery, can enable and facilitate the engagement in value generating activities for business, increase income to households and improve public service quality. Yet, about 840 million people in 2017 (about 11% of global population) had no access to electricity (IEA, 2019a). One sub-target of the SGD 7 is therefore to increasing the share of population with access to electricity and improve infrastructure (UN, 2015).

A recently published tracking report of the SDG 7 specifically addresses energy access issues sub-Saharan Africa (SSA), which is currently the region with the largest energy access deficit in the world. 573 million people, more than a third of the population, lacks access to electricity (World Bank, 2015). In their Africa Energy Outlook, the IEA found that in urban areas, almost three quarters of households have access to electricity whereas in rural areas it is only one quarter of the population (IEA, 2019a). In comparison, the average electrification rates in other developing countries are generally over 90%. Moreover, studies found that in 2016, more than half of lower secondary schools and upper secondary schools, as well as approx. 60% of health care facilities did not have access to electricity (Cronk & Bartram, 2018; UNICEF Institute for Statistics, 2019).

Despite the regional differences and the relative success of some countries to increase their electrification rates in the past years, the problem is prevailing. The difficulty of providing sufficient access to electricity in SSA can be attributed to several factors such as the remoteness of many settlements, poor transportation infrastructure, unfavourable investment environments, low purchasing power of potential consumers and governance issues. A large problem is the rapid population growth, to which the electrification rates of some countries have difficulties to catch up. Combined with the proceeding economic growth, the resulting increase in electricity requirements is expected to lead to a quadrupling of electricity demand, which is expected to be a main driver of global energy demand (World Bank, 2015).

1.2 Problem Definition and Gap in Knowledge

The Annual Energy Outlook (AEO) 2019 found that the currently existing policy frameworks and plans will only reach about a third of currently unmet need. Closing the electricity access gap requires not only an adequate amount of investments but also sound investment strategies supported by a profound understanding and knowledge about requirements, technologies, risks and opportunities backed up by data. (EIA, 2020) Consequently, the body of literature dedicated to assessing the suitability of different electricity generation and distribution technologies for different areas in SSA has seen some growth in recent years. With it, a few existing energy models (such as IMAGE (Integrated Model to Assess the Global Environment) (Dagnachew et al., 2017) and OnSSET (Open Source Spatial Electrification Tool) (Korkovelos et al., 2019; Mentis et al., 2017) have started to direct their attention towards the specific characteristics of developing countries (such as differentiation between commercial and non-commercial energy commodities, traditional, often bio-based, energy resources, the poor-rich divide and the urban-rural divide and the existence of an informal sector (Timilsina, Govinda R; Bhattacharyya, 2009)) as well as more geographically detailed data using GIS (Korkovelos et al., 2019).

One study on this issue is carried out by Dagnachew et al., who aims to identify the most cost-effective distribution technologies per region using a residential electrification model specifically designed for SSA integrated in the IMAGE-TIMER (Targets Image Energy Regional Model) model (Dagnachew et al., 2017, 2018; van Ruijven et al., 2012). This model allows to compare the option of central grid extension and several off-grid solutions on a regional scale, taking into account local characteristics using GIS to model the most cost-optimal technology mix per region (see section 2.3.1 for more detail on the model). Despite relatively fine spatial granularity and detail on electrification options, the approach still has several shortcomings. First, demand is usually projected in a top-down manner, based on simplified macro-level assumptions about population developments. Secondly, most models only capture household demand while demand for public services and productive uses are usually neglected even though these micro-scale productive uses (as opposed to industrial demand, which models treat outside of residential demand projections) have frequently been found to be major consumers of electricity as a wide-spread phenomenon in SSA (Peters et al., 2019; Williams et al., 2015). OnSSET, for example, determines electricity demand by externally setting targets rather than to model it according to actual expected needs and appliance use patterns (Korkovelos et al., 2019). The IEA defines energy access only by regarding the demand from households due to data constraints acknowledging a neglect of productive uses as well as public uses (IEA, 2019b). Currently, TIMER takes into account productive uses but in a top-down, simplified approach by simply adding 20% on top of the projected household demand, which, in turn, is determined by economic development and population dynamics. The barrier for a more accurate, needs-based modelling approach so far was the lack of understanding for how newly connected households can be expected to consume electricity and what determined their demand. This lack of understanding about the actual drivers of electricity demand and the resulting oversimplification of projections can lead to an underestimation or overestimation of electricity demand and risk a waste or shortage of electricity capacity and infrastructure, which can render electrification projects financially unsustainable (Taneja, 2018).

The projections by IMAGE of the impact and cost of different policy scenarios like the subsidization of different technologies, are intended to inform and advice the strategic development policy and investment decisions of the Directorate-General of International Cooperation of the Dutch Ministry of Foreign Affairs. In order to best address this task, the PrElGen (Productive Uses of Electricity Generation model) platform is currently being developed to improve the projection of electricity demand (Falchetta, n.d.). PrElGen is an open-source, bottom-up device-based platform designed to model power demand in SSA based on GIS data and scenario analyses. It is intended to be soft-linked to larger power models such as IMAGE-TIMER or OnSSET. Using a detailed, region-based approach allows to respect national and local differences within SSA. The determinants of electricity demand are ownership and usage patterns of different appliances for different use sectors split for residential, health facilities, schools and activities of emerging local value chains such as improved irrigation systems of agricultural land and the application for productive uses on the micro-scale, especially from emerging processing facilities of locally grown crops (see section 2.3.1 for a more detailed explanation of the PrElGen framework).

The barrier to modelling the demand in such a bottom-up manner so far has been that it is very data intensive and obtaining the kind of detailed and consistent data required to build an accurate model, let alone clearly identify drivers, is problematic for developing countries. In order to bridge the data gap, different simplified projection models for different demand sectors are developed to approximate demand and its determinants where exact data is lacking.

1.3 Research Objective and Research Question

According to the issues outlined in the two previous sections, the gap this thesis aims to fill is threefold. On the highest level, the goal is to contribute to a better electricity access planning to advice strategic development policy planning in SSA by improving the way TIMER is modelling electricity access issues in the region. In order to achieve this, the demand projection is supposed to be refined by the geospatial bottom-up, high-resolution PrElGen framework, which is the second dimension. Within this framework, this thesis is focussing on the electricity demand component for residential-scale productive uses, more specifically for micro-scale enterprises, which are small self-organized income-generating activities undertaken by households (for a more detailed explanation of productive uses see section 2.1), which is the third dimension.[1] Since no representative, encompassing empirical data, let alone census data is available about the electricity consumption of microenterprises (or even their presence), the challenge of this thesis was to find a methodology to estimate this consumption.

The guiding research question for this study is thus:

How can electricity demand for productive uses of households in sub-Saharan Africa in 2030 be modelled?

The research was carried out in two work packages. The first one is aimed at protecting the demand from crop processing facilities as an addition to agricultural activity, the other is aimed at projecting demand from non-agricultural, non-farm businesses. The main research question was therefore split into two sub-questions. According to the challenges faced to answer the sub-questions, several sub-sub-questions were formulated. So far, there is little knowledge about the exact number of EP other than anecdotal evidence on mostly community-level case-studies but it is known that activity varies significantly amongst regions (Nagler & Naudé, 2017). The first step was therefore to find out which predictors can be used to estimate activity. By the same token, it was decided to model the size/performance of EP as a crucial determinant for modelling the actual electricity consumption, which was carried out in a third step. Finally, the anticipated uncertainties associated with building the empirical model require a critical reflection of the expected contribution to the model quality compared to the current top down assumption of adding 20% on top of household demand for productive uses. The sub-questions are structured as follows:

QA. What is the impact of electricity demand from local crop processing residential electricity demand projections?

  1. *What are the different processing operations for the selected crops?*
  2. *How much electricity is required for each step of the processing?*

QB. What is the impact of electricity demand from other non-agricultural microenterprises on residential electricity demand projections?

  1. *What determines the uptake of microenterprises?*
  2. *What determined the performance of microenterprises?*

[1] Only the findings of this third dimension will be presented in this report as to the present date, no final modelling results from IMAGE are available.

3. *What is the electricity demand of microenterprises?*

QC. How well can the adjusted model be expected to project electricity demand for productive uses in 2030?

A graphic representation of the different steps of the research framework that were needed to answer the questions can be found in Figure 1 and Figure 2. Figure 1 is dedicated to the first work package and answering QA and its sub-questions, Figure 2 is dedicated to QB. For the first work package, data on electricity requirements of different crop processing operations were gathered and prepared to be implemented to TIMER. After the first implementation of data, there is a chance that the data needs some revision if modelling outcomes appear implausible. When the final model is ready, the projection results are supposed to be tested for their robustness using a sensitivity analysis. However, as this was outside of the scope of this study, the respective box in the Figure is made transparent.

The second work package was more extensive and consisted of the identification of predictors non-agricultural microenterprise activity (in the following abbreviated as ME or EP) and their electricity requirements based on an explorative approach that consisted of a literature review and a reiterative statistical modelling process. The little existing research on the determinants of uptake and electricity consumption of ME in SSA the anticipated limitations of data availability-imposed uncertainty on the exact choice of model that would be suitable for the given purpose. For this reason, an explorative approach was chosen as methodology from which the final modelling methodology eventually emerged. The reiterative statistical modelling process involved finding available data for all of SSA as *independent variables* in and testing the different variables in different regression models for a range of quality criteria on data sets of households and enterprises which contained information on the *dependent variables*, namely entrepreneurial activity and electricity consumption. The models were required to be conceptualized in a way that would allow them to be linked with each other and with IMAGE and they were optimized for different quality criteria, especially their predictive power. Again, there is the option that revision is necessary if test-runs in IMAGE lead to implausible projection. The main report focusses on the final model result but also indicates relevant alternative modelling methods and reasons they were not feasible in Appendix 13.16 to inform future research. As for the first work package, a sensitivity analysis is to be conducted for the final model.

For the sake of transparency, it is important to note that, despite all attempts to provide the highest levels of validity and reliability, the explorative nature of the approach in combination with limited data availability made it necessary to make many simplifying assumptions and generalizations which pose some noteworthy limitations on the model. These limitations will be explained throughout the report and again summarized in section 9 to answer QC in a final reflection on the findings obtained by the analyses in the context of the different limitations that were encountered.

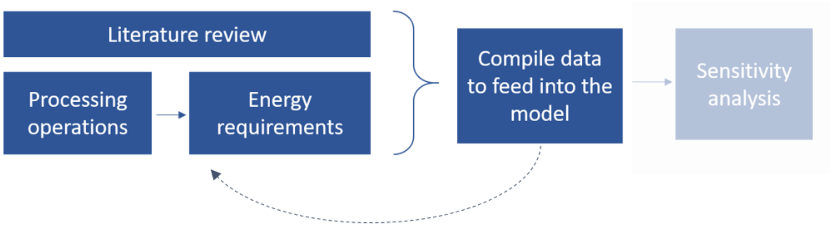

Figure 1. Research framework for crop processing.

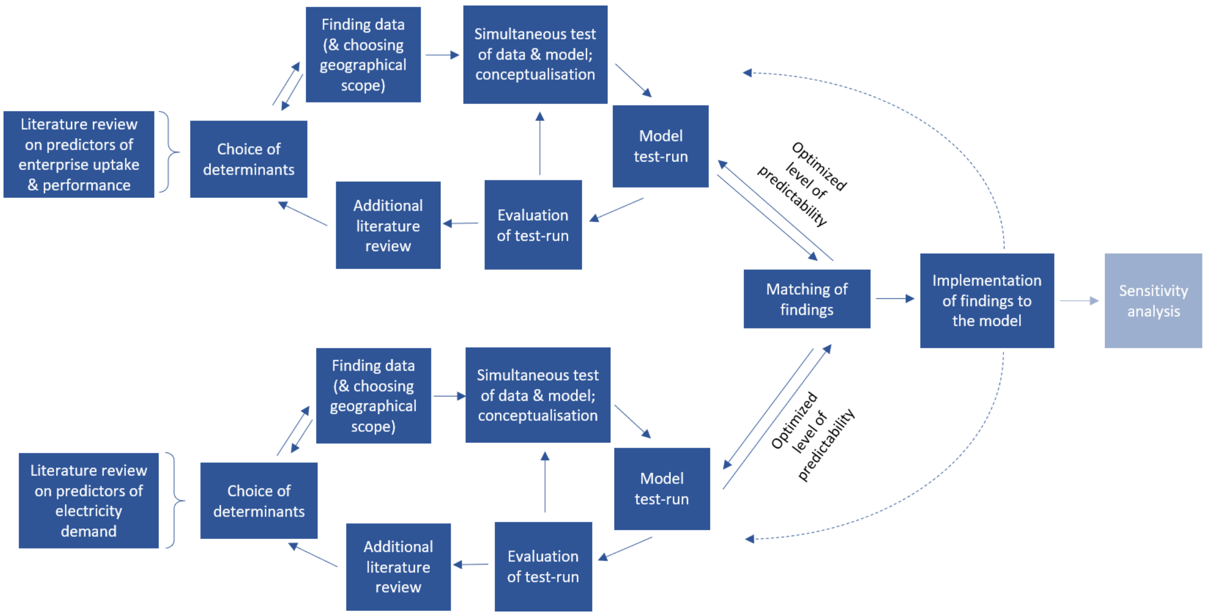

Figure 2. Research framework for other MSMEs.

1.4 Report Outline

More details on the relevant concepts and assumptions made in this research (on microenterprises in SSA, crop processing, productive uses of electricity, energy access modelling and specifically the IMAGE-TIMER model and PrElGen framework) are given in section 2. Section 3 explains the final methodologies used in the research. The findings of the work package on crop processing are presented in section 0. As the research on other productive uses was much more extensive, the following five sections are dedicated to the second work package. Section 5 presents the findings form the literature review, section 6 explains the data sources used, section 7 explains the model conceptualisation and section 8 explains the final models and their performance with respect to several quality criteria. Finally, 9 wraps up the findings of the research in the context of the different limitations, from which section 0 deducts recommendations for future research. Section 11 rounds up the report with an overall conclusion. Additional information on further modelling ideas and deeper investigation of some descriptive statistics are given in the Appendix. To better understand the structure of the report, it should be noted that due to the reiterative approach that was used in the second work package, some decisions at earlier points in the report are based on findings that are only explained alter in the report, in which case, according references are made in order compensate for "breaks" in the linear structure of the report (e.g. the methodology section 3.3 can also be evaluated as a result section.

2 Conceptual Framework

2.1 MSMEs in sub-Saharan Africa

The following section is dedicated to creating a better understanding of the nature of microenterprises (as executor of productive uses) in SSA in general. It will be explained how current literature defines microenterprises, the role they play in the region and which outlooks are provided. Section 2.2 will then take a closer look at their role for electricity access planning and Section 5 will shed light on the predictors of their uptake, performance and electricity consumption as given in current literature.

To this date, there is no global definition of MSMEs due to country specific standards (Berisha & Pula, 2015). Concepts used in developing countries are oftentimes difficult to apply in developing countries. While in developed countries, for example, labour force participation is thought of as dichotomous, the reality for many people in developing countries is that economic tasks (e.g. food growing), economic chores (e.g. collecting firewood) and noneconomic activities (e.g. caring for children), are oftentimes done simultaneously (Fox & Sekkel Gaal, 2008). For the purpose of this study, the definition of the International Finance Corporation (IFC) for micro-scale enterprises will be used, i.e. the focus of this study lies on enterprises which have a headcount below 10 and assets and sales below 100,000 USD p.a. (see *Table 1*).

Table 1. Definition of MSMEs by the IFC. Source: (IFC, 2020).

Scale of Enterprise	Employees	Total assets (USD)	Annual sales (USD)
Micro	< 10	< 100,000	< 100,000
Small	10-49	100,000 - <3,000,000	100,000 - <3,000,000
Medium	50-300	3,000,000 - 15,000,000	3,000,000 - 15,000,000

If not otherwise specified, all terms this report uses, (microenterprise, enterprise, business, ME or EP) refer to this definition. Other frequently used terms in literature on the type of EP at hand are "household enterprise" (HH-EP) or even only "income generating activities", since these activities are indeed so small that they are often carried out inside the home of the entrepreneur. They also often result from a common decision making process of the household, rather than by individual members (Reardon, 2007). This research therefore also focussed mostly on the household (HH) level rather than the individual level. Indeed, few of the enterprises even get any formal kind of registration and make thereby up a large share of the vast informal sector present in SSA (World Bank, 2005). Examples of business activities are small-scale manufacturing such as the processing of food, garments, baskets or furniture, and service activities such as retail, transportation, repair, food services, barbering or street vending (Fox & Sekkel Gaal, 2008). Another useful definition of productive uses is provided by the PRODUSE initiative which aimed at providing evidence on the impact of increased electrification on economic activity. They define productive uses as the additional electricity use of households on top of their own private consumption in order to provide additional (taxable) income (Mayer-Tasch et al., 2013). Examples given for developing countries are agricultural uses (e.g. irrigation), agro-processing (e.g. milling), manufacturing (e.g. carpentry, tailoring, welding and looming) and the service sector (e.g. gastronomy). The authors distinguish productive uses from social uses such as demand from education and health facilities. Opposed to this, some literature argues that anything that enables people to work more productively could count as productive use. This means that health facilities, for example, could also be classified as productive uses because by contributing to the health of people, they also improve their ability to work (Cabraal et al., 2005). For the purpose of this research, it will be stuck to the definition of directly income-related non-farm productive uses since electricity demand from irrigation, education and health facilities are already covered by other components of the PrElGen-framework.

Due to the high share of informality, there are few exact numbers available on the actual size of the ME sector. In 2005, the informal sector in SSA was estimated to account for 90% of all new jobs (CfA, 2005) and 85% of total employment (World Bank, 2005). Fox & Sekkel Gaal (2008) find 80% of labor force in Africa to be self-employed. Analyzing the World Bank household-level surveys, Nagler & Naudé (2017) find an average of 42% of rural HH to be operating an EP but with differing shares in six different countries in SSA (Ethiopia, Malawi, Niger, Nigeria, Tanzania and Uganda; see Figure 3). While most activity is still agricultural, the overall tendency of non-agricultural self-employment is increasing (Haggblade, 2007). There are several reasons for this development.

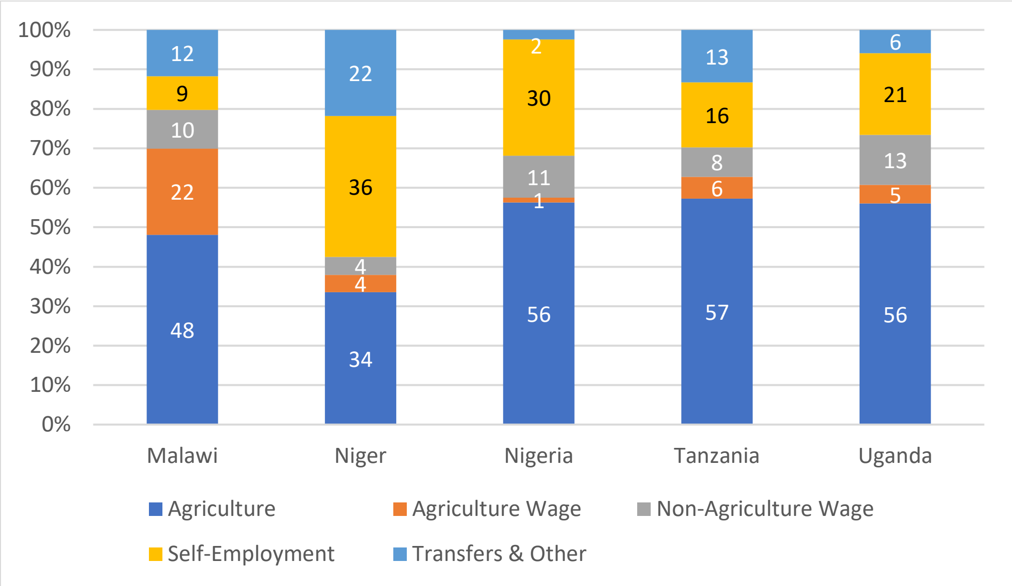

Figure 3. Shares of employment types in five different countries in SSA. Source: (Nagler & Naudé, 2017).

Investigating ten SSA countries (Angola, Botswana, Ethiopia, Ghana, Malawi, Namibia, Nigeria, South Africa, Uganda and Zambia), Herrington & Kelley,2(013) find that the perception of opportunities and availability of skills to start a business of people are much higher than anywhere else in the world. On average 70% of interviewed people expect good opportunities for starting a business in the next six months and 76% perceive that individuals have the necessary skills to start a business with the exception of South Africa (35% and 39%). Accordingly, the share of people who intend to pursue a business in the next three years is very high (53%) as well as the share of early stage entrepreneurial activity (28%). Conversely, the share of people with established businesses varies between 5-15%. The largest difference between early stage and established business activity is mostly found in factor-driven economies and countries with a low GDP per capita. Despite this indication for high rates of discontinuity, SSA also has the lowest levels of fear of failure.

However, most of the early entrepreneurial activity is needs-driven rather than aiming for profit maximization and growth. The differentiation between needs-and opportunity-driven businesses is common literature on MEs in developing countries (Kooijnman-van Dijk, 2008) and also plays an important role in the present study. Needs-driven or survivalist EP are “pushed” into business as an attempt to escape or avoid slipping deeper into poverty. Their intention is oftentimes to diversify their income source to mitigate risk from agricultural income sources, which are becoming increasingly uncertain due to rising populations, associated landlessness, more extreme weather events due to climate change and other agriculture-related shocks (Szirmai et al., 2013). Entrepreneurs of push-businesses are usually less prepared before starting business and focus on easy-to-enter businesses such as trade. Opportunity-based EP are pulled into business by the prospect of developing and growing a business. They seek profit maximization and are generally better prepared to ensure availability of adequate resources (Reardon, 2007). This duality makes clear that the mere level of EP activity is not an indication which can be evaluated by itself as “good” or “bad” because it can be both, a symptom of poverty, stagnating agricultural outputs and lack of alternatives, as well as a thriving and dynamic economic environment (Haggblade et al., 2007). In the context of this research, this means that some predictors might have different effects on the presence of different types of businesses (see section 5).

In this context, there is also an ongoing search to determine the actual potential of MEs to alleviate poverty, absorb the growing labor force and further economic development. To this point, the evidence is mixed. The contribution of MEs to development, especially in terms of offering great potential for employment, are generally acknowledged (Okpara & Wynn, 2007) and MEs have received policy support in different countries (Herrington & Kelley, 2013; Urban & Naidoo, 2012). However, the general finding concerning poverty alleviation is that MEs seem to have the ability to keep people from slipping further into poverty (Fox & Sekkel Gaal, 2008; Szirmai et al., 2013), but their ability to lift people out of poverty depends on the favorability of circumstances (Haggblade et al., 2007). Some studies also voice concerns and potential risks of a large ME sector in developing countries. Little et al. (1987) question their efficiency, Hallberg (2000) see their vast occurrence as a symptom of market imperfections rather than “intrinsic” economic efficiency and Lipton (1977) and Southall (1980) question the overall ability of rural areas to stimulate growth especially where unequal wealth distributions allow richer businesses to “siphon surpluses”. The importance of efficiency becomes especially salient in the international context where efficiently producing countries trump low-priced products local EP cannot compete with (Fox & Sekkel Gaal, 2008). In their study of 96 micro-, small and medium domestic shoe producers in Addis Ababa, Gebre-Egziabher (2009) reports that as a consequence of Chinese competition, 28% of EP were forced into bankruptcy, and another 32% downsized activity. The average size of microenterprises fell from 7 to 4.8 employees, and of SMEs, from 41 to 17.

Concerning productivity, Aterido & Hallward-Driemeier (2010) indeed found evidence of lower value-added power per worker in microenterprises in SSA as compared to larger EP as well as compared to ME in other low-income regions. In the context of this strong concentration of employment in generally more labor-intensive ME, they voice concerns about productivity growth as well as the efficiency of resource allocation. However, rather than making a clear statement about the desirability of employment in ME, they emphasize the relevance of policy makers to be aware of effects policies can have on different scales of EP. Improved conditions and expanded employment opportunities of more labor-efficient large scale enterprises might not immediately translate into a decrease in ME employment due to other barriers such as lack of qualification. The aim might rather be, at least in a first step, to address constraints which keep ME from operating more productively rather than trying to shift employment to larger EP. The African Entrepreneurship Report 2012 concludes “While entrepreneurship may not be a panacea, it can most certainly form part of the solution [to high poverty, unacceptable levels of unemployment]” (Herrington & Kelley, 2013).

Irrespective of the normative stands about the desirability of ME, Fox & Sekkel Gaal (2008) find that “even under the best of scenarios” wage and salary jobs will not be able to absorb the majority of the strongly growing labor force until 2030, especially in urban areas.

Against this background, rather than engaging in the debate about desirability of EP, this report acknowledges the ample evidence that micro-scale entrepreneurship will continue to form a major share of employment and probably even gain relevance. It is further acknowledged that there is a noteworthy heterogeneity amongst ME and multitude of drivers which reinforce ME activity across SSA. These findings emphasize the relevance of this research and are important considerations to keep in mind for adequate representation of ME in the model and the interpretation of the projection results.

2.1.1 Crop Processing

Among the different types of MEs, increased and improved crop processing is especially expected to become more relevant in rural SSA. As opposed to economies of other developing countries, the main sector of employment in SSA is still agriculture while only adding a relatively small share of value to the total GDP. Nevertheless, it is usually assumed in literature that agriculture will continue to be an important sector in the future of SSA (IEA, 2019a). Current processing techniques such as the milling of millet, rice, maize and cassava to use for gruels and porridges is done often done in traditional, laborious mortar and pestle fashion (Ajala & Gana, 2015). Electrification of these processed could immensely improve productiveness and quality, which can not only improve income but also alleviate people form this time- and labor-intensive work. The issue of quality has been found by Ajala & Gana (2015) to keep people from consuming their locally grown products. They state that Nigeria had the potential to be self-sufficient in their rice consumption, but people prefer to eat imported rice because of the poor quality of rice and occurrence of stone resulting from traditional processing techniques.

At the moment, there is still a lot of subsistence farming and a large share of crops grown for trade are usually sold without much previous processing. Establishing regional agricultural value chains such as crop processing activities would be an opportunity to increase the local value creation, the export of higher-quality end-products and ultimately, local wealth (Terdoo & Feola, 2016; UNCTAD, 2019). Some examples such as a case study on post-harvest cocoa processing in Uganda revealed a general interest amongst smallholders as they acknowledged the potential to improve product quality, increase the retail prices and household incomes (Jones et al., 2011). Due to these beneficial effects of local crop processing measures, it can be expected that in the future, more processing activities will take place in rural areas. Depending on the technology, this could be a substantial additional source of electricity demand which was so far neglected in modelling approaches (Mapako & Mbewe, 2013).

2.2 The Meaning of Access to Electricity for Productive Uses

While the main focus of this study is the specific impact of electricity on productive uses, it is still important to first understand the broader context in which access to electricity is defined in literature and development projects and the premises on which electrification projects are justified. The motivation behind most current electrification projects and research is the generally assumed contribution of electricity to economic development. This idea is often supported with illustrations of

S-shaped correlation curves of GDPs or HDIs and electrification rates, thereby suggesting a causal relationship (for an example see Figure 24 and Figure 25 in Appendix 13.1).

These studies tell the narrative that a productivity improvement from using electrical or other forms of energy as compared to manual labor can generate more wealth by improving productivity. As countries industrialize, they eventually achieve better energy efficiency and shift to less energy-intensive sectors, thereby narrowing the curve. However, finding statistical evidence is a methodologically challenging task, which only few studies have undertaken (Mayer-Tasch et al., 2013). The difficulties in assessing electricity access are the intermediary character of electricity (and therefore non-linear relationship to the performance indicator) and the non-binary character of electricity supply due to the prevalent deficiencies in developing countries. In other words, only having a physical connection to a power source does not mean that electricity is always available. The World Bank (2015) provides a multi-tier framework in which they form multi-tier matrices for different types of energy supply and consumption groups by differentiating access on different dimensions such as peak capacity, duration of availability, reliability etc. A literature review conducted for the PRODUSE report on macro-level impacts of electrification on productivity, growth and poverty alleviation finds that, while most studies do find statistically sound evidence for these correlations in the cases they analyzed, there is also some noteworthy contradiction from other cases (Mayer-Tasch et al., 2013). They conclude that the effect depends highly on country and context and that electricity can be a strong enabler but not a sufficient condition for macro-level developments. Concerning poverty reduction, for example, other infrastructure improvements, e.g. of roads and water supply, are often found to be more relevant than electricity. Overall, the evidence for benefits from electrification is stronger when looking at specific use cases such as (cooling) applications for health facilities, clean cooking, telecommunication etc. and more specific indicators of the potential benefits, such as health, safety and wellbeing.

An interesting question in the present study is in how far electricity contributes to the uptake and performance of household enterprises (see Appendix 13.18 for an explanation of the acceleration effect in consumption caused if EP uptake is affected by electrification). While current findings from analyses in literature on the relationship between entrepreneurial activity and electricity consumption will be investigated in section 5, a brief insight into the theory on how electricity can lead to improved productivity is provided here.

Potential benefits of access to electricity are efficiency improvements, extension of variety and quality of services and goods offered locally, resulting in access to new demand of the existing customer base but also of customers from outside the region (Booth et al., 2018). Electrified areas can also attract new residents and businesses from non-electrified regions. A useful visualization for different pathways was developed by the PRODUSE report (Mayer-Tasch et al., 2013).

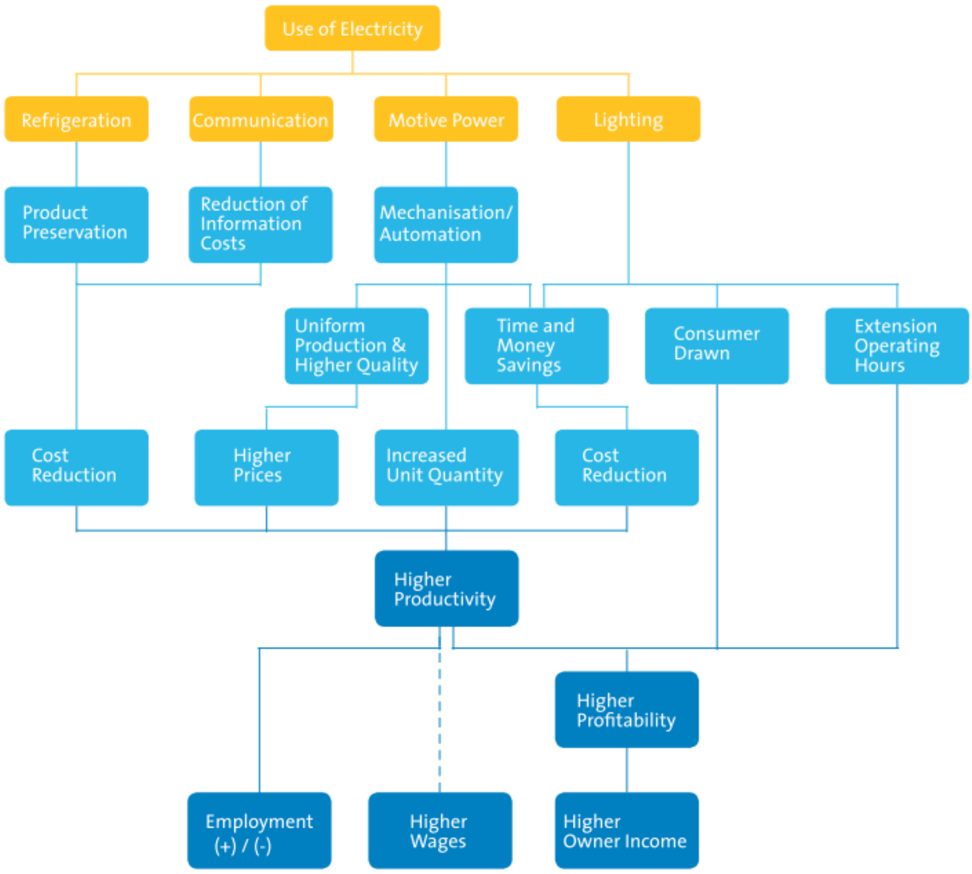

Figure 4. Pathways from Electricity to Income Generation. Source: Mayer-Tasch et al. (2013).

But the benefit of productive uses is not only on the side of the ME. In the best-case, there is a win-win scenario in which the enterprises can use electricity to improve their businesses thereby contributing to the profitability of electrification to utilities as well, helping them to make the electrification economically sustainable. This is hoped to be one solution to break the vicious cycle where poverty in an area leads to lack of purchasing power of households, which, in turn, keeps electricity providers from establishing connections there that could be used to generate more wealth (Booth et al., 2018; Brew-Hammond, 2010). Even though their activities might seem small, the previously mentioned prevalence of ME in SSA make them a major potential consumer of electricity (Williams & Jaramillo, 2018). Moreover, considering the low consumption levels per household in some regions (65 kWh/year), each additional kWh can increase the relative demand tremendously (Dagnachew et al., 2017). Commercial consumers also usually pay higher tariffs than households. Blodgett et al. (2016) find that in their analysed micro-grids, the top 10% of consumers are businesses who use energy-intense appliances and consume 50% of electricity and provide 40% of revenues to the electricity provider.

However, even to the extent to which electricity access can help improve businesses to connected EP, some studies point out important potential pitfalls. In some circumstances, electricity access can also widen the gap between wealthy EP who can afford to use electricity for their profit maximization and thereby take away some of the consumer base of poorer or more remote EP with new, better or cheaper products (Harsdorff & Bamanyaki, 2009). As long as businesses only operate within their local environment where a generally low efficiency is the standard, businesses might not even perceive a disadvantage due to lack of electricity.

This might also be a reason why most literature finds the relationship between electricity supply and business operation to be not as straightforward as suggested by the theory but rather, as for the macro-effects, to be a matter of research methodology and context. The availability of electricity is often found to be a necessity to access certain income generating activities and grow businesses, but it is not a sufficient factor for success (Wilcox et al., 2015). Oftentimes, connection rates in newly connected areas remain rather low, as do appliance use and consumption levels (Mayer-Tasch et al., 2013). Barriers are usually too high costs for connection and consumption, which are amongst the highest in the word in SSA, as well as high investment cost for electrical appliances (Golumbeanu & Barnes, 2013; Lenz et al., 2017). Additionally, quality of supply in terms of low or fluctuating voltage and outages can result in workflow interruptions and damage of equipment. Due to the lack of supply, many HH and EP rely on alternative electricity sources such as diesel generators. Foster & Steinbuks (2009) find own-generated electricity to make up a share of up to 25% of all installed capacity in some SSA countries. These sources are usually more expensive to operate, security of supply might also be limited in some areas and they tend to have some serious environmental and sanitary effects (Neelsen & Peters, 2011; Scott et al., 2014). Within the inconsistent empirical findings on the benefits of electrification to business operation, the high generator ownership surely is a sign of the high unmet electricity demand by EP.

2.3 Modelling Access to Electricity

In order to understand the current state of research on electricity access planning, the following section gives a brief insight to the most recent findings, assumptions and limitations.

To begin, it is important to differentiate between electrification access planning on a project-scale basis and for strategic policy planning. Both are equally important but are faced with very different challenges. Project planning is dealing with question of site selection and means to promote and stimulate connectivity and with the involvement of local stakeholders, more detailed information of target region characteristics is available. For strategic policy planning, only much more aggregated macro-level data tends to be available, however, the nature of questions are also different. In the case of the present project, it is mostly of interest to know the order of magnitude for electrification cost and favorable technologies for the regions of Central, Wester, Eastern and Southern Africa (with exclusion of South Africa). Rather than to know details about regional specifications, the modelling is based on the factors that are generalizable across a broader geographic area. Of course, a more detailed knowledge is always desirable (hence the bottom-up modelling approach of the present study) but it is much more difficult to achieve on such a large scale. As planning gets more concrete, the spatial scope is narrowed down, and the focus can be put more on locally specific traits and needs which influence electricity demand rather than their generalizability.

As previously mentioned, the PrElGen framework is aiming at improving the projection of electricity demand in IMAGE to analyses different policy scenarios to inform strategic policy planning. The model can provide a cost-optimized technology mix for universal access as well as model impact and cost of different policy scenarios. So far, Dagnachew et al. (2017) found that the baseline developments under the existing policies only do not lead to universal electricity access by 2030. Especially in rural areas access rates remain low. From developing a scenario for universal access by 2030, the authors found that the preferred electrification technology to be strongly influenced by the targeted level of consumption. While off-grid solutions showed to be preferable for low levels of consumption, extending the central grid was more economical for high levels of consumption. At their baseline level of consumption (between 365 and 1250 kWh/household p.a.), 85% of the population with new access to electricity could be connected to the central grid, even 95% for high levels of consumption (between 1250 and 3000 kWh/household p.a. or more) due to economies of scale. At lowest levels (around 4.5 kWh/household p.a.), 65% of population could be connected to decentralized systems, mostly stand-alone systems (50%). The difference in technology requirements also has strong implications for the investment requirements: the authors found the cumulated investment needed to enable universal access to electricity in SSA from 2010 to 2030 to be 22 billion USD for low levels of consumption and 2.5 trillion USD for high levels of consumption. For the baseline, demand was modelled by a bottom-up calculation based on assumed appliance ownership at a given GDP per capita level, for the other scenarios, electricity demand levels were based on a top-down uniform assumption for every newly connected household based on the Multi-Tier Framework of the SE4ALL Initiative (World Bank, 2015). The AEO 2019 projected a similar need for investment for full access in 2030 but assume a high share of over 50% of decentralized systems for their baseline scenario. Another study found solar home systems to be the fastest option in most cases (about 40 to 70% due to model uncertainty) but also stresses the importance of grid-based systems due to their broader technical reach, i.e. for high demand industry use (Bertheau et al., 2017). Korkovelos et al. (2019) used the OnSSet model and the Multi-Tier Framework by the World Bank (2015) to project energy access in both rural and urban areas in Malawi in 2030 and came to similar conclusions as Dagnachew et al. (2017) concerning the strong impact of the projected demand level on the most cost-optimal technology mix. The diversity of optimal technology mixes these projections provide and their sensitivity to different demand levels emphasize the need for more accurate demand projections.

2.3.1 IMAGE-TIMER and the PrElGen framework

In order to better understand the context of the present research, i.e. how the findings of the thesis fit into the PrElGen framework and the IMAGE model and finally how the findings of the IMAGE model are generated to comprehend how they can be interpreted, it is worth taking a closer look at the properties and logic of the IMAGE model and the PrElGen framework.

IMAGE 3.0 is an Integrated Model to Assess the Global Environment by PBL with TIMER as its energy component. The purpose of it is the large-scale, long-term assessment of the different global environmental impacts of human development (see Figure 27 in Appendix 13.2 for an illustration of the components in IMAGE). It covers individual characteristics of 26 regions. TIMER, more specifically, is designed to allow modelling long-term trends in energy consumption and generation and associated greenhouse gas (GHG) emissions by providing a bottom-up least cost-optimization modelling of electrification pathways to the IMAGE framework as well as modelling the implications of different policy scenarios.

The exact model characteristics of the electricity access model in TIMER used for SSA are outlined in van Ruijven et al. (2012), Daioglou et al. (2012) and Dagnachew et al. (2017) and are summarized in Figure 28 in Appendix 13.2. Its main purpose is to make projections about future electricity access rates, choice of electrification technologies and associated investment needs either under given scenario assumptions or for achieving a specific electricity rate.[2] The model is based on a temporal unit of years and a spatial resolution of 0.5° x 0.5° grid cells with information on cost of power generation technologies, population density, household electricity demand, cost of transmission and distribution, technical potentials of renewable energy sources, and distance from existing power lines. This information can be used to model the optimal least-cost electricity generation and distribution

[2] The decision tree can be seen in Figure 26 in Appendix 13.2.

technologies per region to provide universal electricity access. The distribution technologies include the option for grid expansion as well as eight off-grid electrification technologies. The off-grid technologies are a selection of mini-grid options on one side and a few stand-alone systems on the other side. Mini-grids have enough capacity for a single community or small town and are powered by either a diesel generator, solar photovoltaic (PV), wind power, a small hydro plant, a hybrid diesel-PV generator or a hybrid wind-diesel generator. These systems have the potential to be connected to the main grid in the future. Stand-alone systems are designed to provide electricity to individual households and are represented in the model by solar-home systems (SHS) and diesel generators. Furthermore, the model takes into account several local characteristics such as the distance from existing power lines, population density, household electricity consumption, resource availability and the price of the different technologies. These factors determine the cost per kWh of electricity. The split between rural and urban is adapted from the definition on the UN (PBL, 2014), who uses country specific definitions (UN, 2017). The projections will be based on the moderate socio-economic scenario (SSP2) within a variety of available scenarios embedded in IMAGE (see Appendix 13.2.1 for definitions).

As indicated in section 1.3, the PrElGen framework is being developed in order to provide a more detailed, multi-sectoral bottom-up assessment of electricity demand, which could then be linked to existing models such as IMAGE, EU-JRC or OnSSET (Open-Source Spatial Electrification Tool). It should further allow to compare the required investments of increased electrification with its potential economic benefits. The data is obtained from openly available sources and contains information such as the ownership and usage patterns of appliances for households, schools, and healthcare facilities, as well as location, extent, and type of croplands, irrigation water gaps, infrastructure and market proximity. Demand profiles are created stochastically using the RAMP (Remote-Areas Multi-energy systems load Profiles) model, which is suited for developing countries, where little information is available on actual usage profiles (Lombardi et al., 2019).The spatial functional unit of the model are heterogenous polygons based on population distribution patterns.

Figure 5 shows the different demand sectors considered by the PrElGen framework and how they link to SDGs: residential, social welfare in the form of healthcare facilities and schools and the productive use for irrigation. Appendix 13.2 shows the more detailed schematic workflow for each of these demand sources (Figure 29, Figure 30 and Figure 31). Due to data availability reasons, the framework has so far mainly been applied to Kenya. The modelling undertaken in this thesis for the demand from MSMEs and crop processing will add to the productive use sector of the framework.

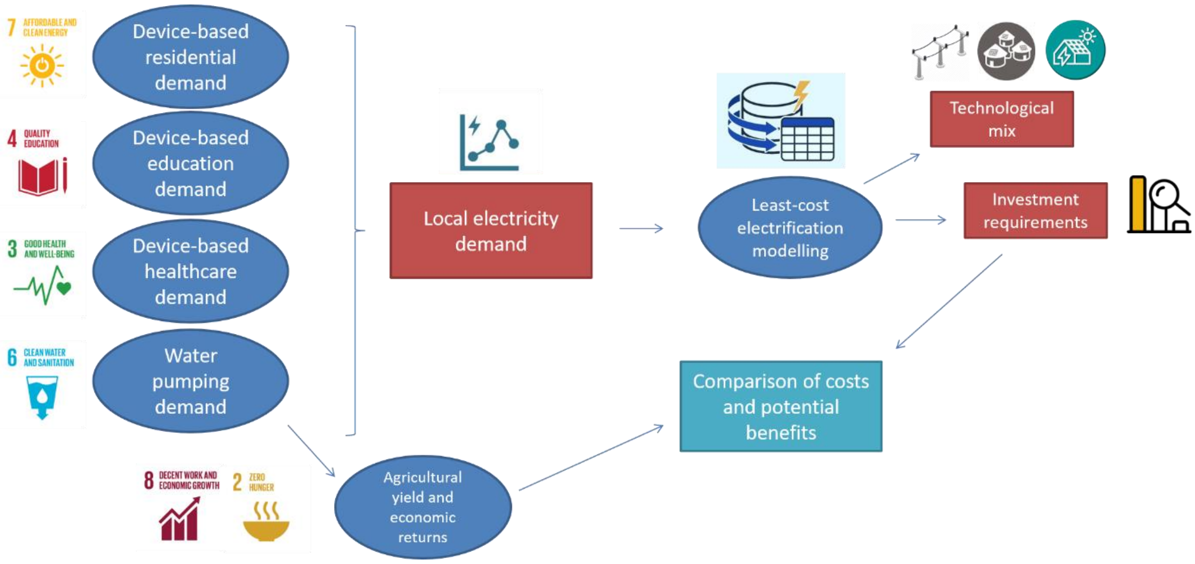

Figure 5. PrElGen Framework. Source: (Falchetta, n.d.).

3 Methodology

3.1 General modelling context of IMAGE and TIMER

As the research question is twofold, there are also two distinct lines of research which will be outlined in the following sections. The findings of the methodological steps are the main expected output of this research. However, once the models are finished and all required data has been collected, this information, together with the data from the other PrElGen models, is supposed to be implemented in the IMAGE model to project the electricity demand in across SSA until 2030. The next step is then to compare in how far the bottom-up results differ from previous model runs in Dagnachew et al. (2017, 2018, 2020). The new findings are then used to derive information about potential future costs and technologies of electrification across SSA to inform and advice development policy strategies to the ministry of foreign affairs. However, within the scope only the final model concepts and projection preparation are presented while the projection results are outside of the scope.

3.2 Crop Processing

3.2.1 Literature Review

The research on crop processing consisted, to large extend, of an extensive literature review of both academic and grey literature. The goal was to find out which processing operations exist for the most important crops in SSA and what the respective electricity needs are. IMAGE already contains information on which crops grow where and the potential yield. *Table 2* shows the selected crops for this study based on their presence across the entire region of SSA. The "Crop category" column shows the crops given in IMAGE, the column "Crops represented" is the explanation of what is summarized in the respective categories.

Table 2. Relevant crops for the research on crop processing.

Crop category	Crops represented
Wheat (spring/winter)	Temperate cereals (wheat, rye, oats, barley, triticale)
Rice	Rice
Maize	Maize
Millet	Tropical cereals (millet, sorghum)
Field peas	Pulses
Sugar beet	Temperate roots and tubers
Cassava	Tropical roots and tubers
Sunflower	Sunflower
Soybean	Soybean
Groundnut	Groundnut
Rapeseed	Rapeseed
Sugar cane	Sugar cane

The additional information required was thus what common products and according processing steps exist for which type of crop, what processing steps are most likely to take place locally and the demand of electricity for each processing operation per unit of crop yielded. Another selection criterion was that only those processing operations should be considered which are feasible on a micro-enterprise scale, i.e. requiring no more than five people for operation and being profitable for a micro-scale EP with maximum sales of 100,000 USD per year. Most literature sources did not provide all required information, so a main task was to find a strategy to prepare the sparsely available information to extract the information in demand. A full explanation of assumptions is be given in Appendix 13.3. the most serious limitations will also be addressed in section 9. The findings of the literature review are given in section 4.

Sources were sorted according to crop and for each source a table was created capturing the different processing steps, their electricity requirements per unit of crop processed and an adjustment of the data in a way in which It can be implemented into the model. To match the data into IMAGE, the electricity consumption per crop had to standardized to kWh/kg yield.

3.3 Non-agricultural MEs

The approach to answer the research questions on the emergence of MEs is mostly of explorative character as the goal was to develop a methodology from scratch to integrate into the existing IMAGE model. Some important boundaries of the research had to be determined only in the course of the project, such as the exact geographic scope, data sources and the exact type of mathematical modelling approach because these factors depended on the findings in literature, data availability and on each other. Due to their interdependence, these decisions were made and reassessed in the re-iterative approach outlined in the research framework in *Figure 2*. In this sense, the development of a methodology was part of the study objective, this section can be understood as the answer to the main research question. However, in order to keep a linear narrative for this report, only the final methodology will be presented and further information on the modelling choices are given in the Appendix 13.17 and 13.16.

3.3.1 Literature Review

As illustrated in the research framework, the research began with an extensive literature review to find predictors associated with the presence of MEs in SSA. This section will be referred o in the following as “uptake” model, even though the existence of an EP is, of course, not only the result of uptake but also the ability to survive. A further literature review was aimed at identifying the electricity demand of these enterprises. As It was found to be beneficial for the accuracy of the electricity consumption model to also model sales of an enterprise, the literature review also focused on the predictors of sales and performance metrics in general, which are, however, closely linked to the predictors for EP uptake (for more details on the model choice see section 6.3.2). The predictors of electricity consumption were mainly found in studies which considered the decision to connect as dependent variable and studies which considered the level of consumption, which were considered equal.

The literature review was focussed on SSA countries but some exceptions were made if deemed relevant.[3] However, rather than spanning the whole of SSA, information was usually provided only for single regions and in the form of qualitative, non-representative case studies instead of regionally (meaning SSA), nationally, or even state level representative samples.

3.3.2 Model conceptualisation

At the junction of findings from literature, data availability and data analysis, a theoretical concept emerged (see section 7). For each category, uptake, performance and consumption of electricity, it captures the main interactions of predictors that were found. Its purpose was used to structure concepts and data, clarify interactions and thereby offer guidance to the specific issues and questions that should be addressed by the models. In this report it serves as a visualization of the main interactions addressed in throughout the report and should help to understand the ideas behind the modelling choices. Naturally, the concepts were frequently adjusted along the research process so only the final version is given with some indications where noteworthy simplifications had to be made.

3.3.3 Data analysis and modelling

The challenge of finding a model was to get from the available data in IMAGE, namely the number of households per grid cell in 2030, and some other, so far unknown, data to the electricity consumption of microenterprises.

In attempting to fill the gap with reasonable, feasible steps, several modelling approaches were tested and discarded, and modelling choices changed frequently from early attempts that followed first insights about drivers gained from literature and the use of first data sources. As new data was found, new insights about predictors were obtained from data analyses and new modelling ideas were inspired as well as limitations encountered until eventually, the final model got distilled. Figure 6 outlines modelling steps undertaken to get from the available raw data to the total electricity consumption per grid cell. This section will explain the final mathematical models (white boxes) and how they are connected with each other and with IMAGE, while further, important but discarded ideas are addressed in sections 13.16 and 13.17.

[3] Exceptions are (Kooijnman-van Dijk, 2008; Rao, 2013), who focus on India.

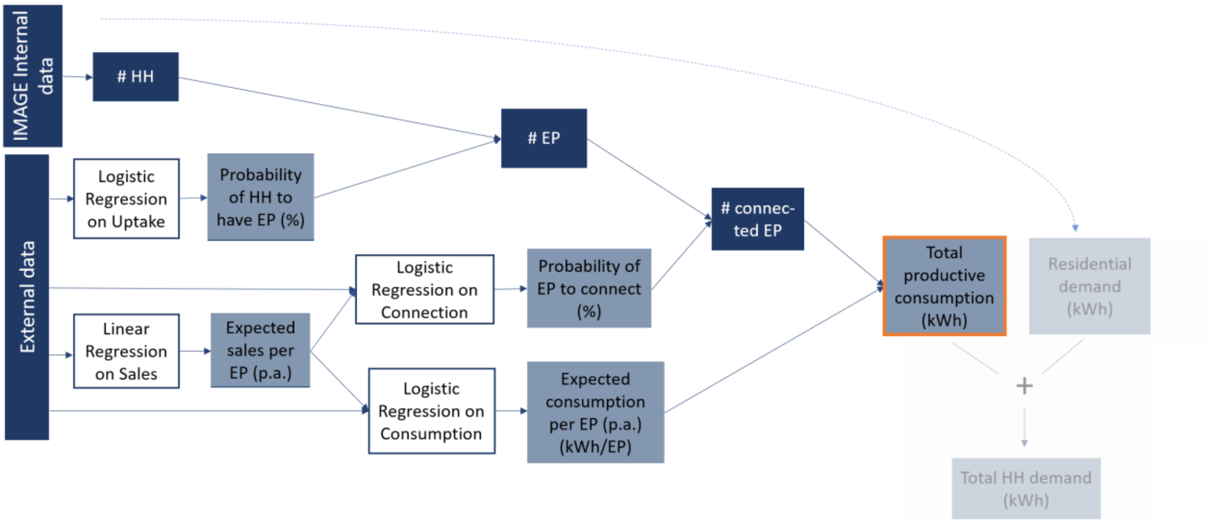

Figure 6. Outline of the modelling methodology. Values are given per grid cell per year.

The final model consists of four distinct regression functions: a binomial logit regression on the probability of a household to be engaged in entrepreneurial activity ("uptake model"), another binomial logit regression on the portability of an EP to have access to electricity ("connection model"), a multiple linear regression on the total annual sales of an enterprise ("sales model") and another multiple linear regression on the electricity consumption of an enterprise ("consumption model").

The regressions are linear functions in which independent variables are multiplicated with coefficients and added up to determine the value of the dependent variable. The coefficients are determined using a least-squares-based algorithm embedded in the modelling software; for this research SPSS 27 was used. Trying out different methods, the ENTER method[4] turned out to lead to the most accurate and transparent results in all models. Before modelling, the independent variables were tested for their linear correlation with the dependent variables using the Pearson Correlation test. More information on the variable choice, quality criteria and interpretations of the final covariates are given for each model in section 8.

Georeferenced WB household survey data with information on household-level entrepreneurship served as a database to test various georeferenced variables from within the survey and several other sources for their power to predict enterprise uptake. The binomial logistic regression function calculates logits, i.e. the natural logarithm of the odds of a certain even to take place, here the event is entrepreneurial activity of a household. The logistic regression function is given by

$$\ell_U = \log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum_j \beta_j X_i + \sum_k \beta_k D_k \quad , \qquad \text{Eq. 1}$$

where ℓ_U is the logit term of the probability $\mathrm{P_i}$ that a given HH (i) will operate an EP, X_i is a continuous variable, D_k stands for a dummy variable of a categorical variable k, β_j and β_k are the respective coefficients of the variables and β_0 is a constant term or intercept. The logistic regression equation can

[4] In the enter method, all variables are added to the model in one step as opposed to stepwise procedures, whereby variables are either entered or removed from the model sorted by their partial correlations.

be used to calculate the probability of the households in a grid cell to pick up a business by inserting the according data for each grid cell and solve the equation for $\mathrm{P_i}$.

$$P_i = \frac{e^{\ell_U}}{e^{\ell_U} + 1} \qquad \text{Eq. 2}$$

As data is given on a grid cell rather than on a HH level, every HH within a grid cell (g) will have the same probability for enterprise uptake, in other words: the household i is representative of all households in its grid cell.

$$P_{i,g} = P_{1,g} = P_{2,g} = P_{3,g} \qquad \text{Eq. 3}$$

Applying the law of large numbers, it will be assumed that the probability of uptake for HH i in a grid cell is equal to the share of HH with EP for that grid cell, i.e. it will give the number of HH with EP (hh_{ep}) per total number of HH (hh_{total}).

$$\mathrm{P_i} = \frac{\mathrm{hh_{[ep,g\}}}}{\mathrm{hh_{total,g}}} \qquad \text{Eq. 4}$$

Furthermore, the number of households with EP is assumed to be equal to the number of EP, even though some HH might have several EP.

$$hh_{ep,g} = ep_g \qquad \text{Eq. 5}$$

The calculated probability obtained from the logit provided by the regression function can therefore be multiplied with the number of HH to model the number of enterprises in a grid cell.

$$ep_{total,g} = hh_{total,g} * P_i \qquad \text{Eq. 6}$$

The initial idea was that the second and final step would be to model the electricity consumption of the present enterprises, but not all enterprises can be expected to even get an electrical connection. Therefore, the next step in the projection was to identify what determines if an EP even has an electrical connection or not. The model was again based on a logistic regression

$$\ell_C = \log\left(\frac{Q_m}{1 - Q_m}\right) = \beta_0 + \sum_j \beta_j X_i + \sum_k \beta_k D_k \quad , \qquad \text{Eq. 7}$$

where ℓ_C is the logit term of the probability $\mathrm{Q_m}$ that a given EP (m) will have an electrical connection. Again, the according data can be inserted, and the equation can be solved for the probability of an EP to have an electrical connection:

$$Q_m = \frac{e^{\ell_C}}{e^{\ell_C} + 1} \qquad \text{Eq. 8}$$

As the data is given on grid level again instead of EP-level and EP m stands for all EP in a grid cell g, it will be assumed that the probability of connecting will be equal to the share of EP with a connection (ep_c) as part of the total number of EP (ep_{total}) in a grid cell.

$$Q_{m,g} = Q_{1,g} = Q_{2,g} = Q_{3,g} \qquad \text{Eq. 9}$$

$$Q_m = \frac{ep_{[c,\mathrm{g}\}}}{ep_{\mathrm{total,g}}} \qquad \text{Eq. 10}$$

Therefore, the number of EP with electrical connection can be calculated by multiplying Q_m with the total number of EP per grid cell.

$$ep_{c,g} = ep_{total,g} * Q_m \qquad \text{Eq. 11}$$

Next, the actual level of consumption needs to be modelled. As the sales turned out to be an essential predictor for the electricity consumption, the first step was to build a linear regression model for the sales of an enterprise. This again required data and a regression function. In this case, the World Bank EP surveys were used as a sample to test different variables for their ability to predict the annual sales of an enterprise. The linear regression function is given by

$$s_m = \beta_0 + \sum_j \beta_j X_i + \sum_k \beta_k D_k + \varepsilon_j \quad , \qquad \text{Eq. 12}$$

where s_m is the value of the dependent variable, here the total annual sales in USD, of an EP m. Inserting the data for each grid cell, this will give the expected sales of each enterprise in the grid cell. The calculated sales can then be inserted into the connection logistic regression function and the linear electrical consumption regression function.

The consumption is modelled again with a multiple linear regression function using the EP Surveys as database and selected predictor variables as well as the previously computed sales:

$$el_m = \beta_0 + \sum_j \beta_j X_i + \sum_k \beta_k D_k + \varepsilon_j \quad , \qquad \text{Eq. 13}$$

where el_m is the value of the dependent variable, here the annual electricity consumption in kWh, of an EP m. The values for electricity consumption can then be multiplied with the number of connected EP in the grid cell to calculate the total expected electricity consumption in kWh per grid cell.

$$el_g = el_m * ep_{c,g} \qquad \text{Eq. 14}$$

3.4 Assessing the Quality of the Projection Results

For building the model, the collection and preparation of input data as well as the programming in IMAGE was done on the side of PBL but the examination for mistakes and fixing of data and programming code were a shared effort.

The actual future uptake, performance and electricity consumption of EP in 2030 is, of course, impossible to know at this point considering that even for present EP, there is little knowledge about general electricity consumption patterns. This condition makes it also impossible to accurately answer the third research question and state the exact extent to which the predictive quality of the demand projection in IMAGE could be improved by the present model. There are some ways, however, to assess in how far the model can be *expected* to be improve, and it can be discussed in how far projection results seem plausible.

While a diligent, critical reflection of limitations encountered in the building process of the model is one central step to gain an idea about the predictive power of the model, another important component is to examine the actual projection results to evaluate in how far the they can be deemed sensible and reliable. As often in statistics, a large share of this evaluation is based on subjective perception but there are further, "hard" strategies to quantify the potential fit of the projection. This section provides some ideas for plausibility checks to ensure the projections are not too far off track.

First, a similar approach can be taken as it was also taken to evaluate the calculated expected values by the models as done in section 8, i.e. the projected results can be compared with values (mean, median, minimum maximum, etc.) of the sample and literature (see appendix 13.7.1). Naturally, they can be expected to deviate to some extent but should not be too different in their order of magnitude. Here, effects of the biases addressed in section 9 should be paid special attention to. To identify the source of possible deviations from the sample, the combined effect of the independent variables can be decomposed and for each variable, the range and frequency distributions of values of the input data can be compared with the sample data that was used to build the models.

As the previous assumption of additional electricity consumption per HH was simply assumed to be 20% on top of the EP demand, it should be compared how much more the new bottom-up model projections differ from this assumption. By looking at a more disaggregate scale and examining consumption levels of average households and average EPs of different regions it can be explored how the values for projected consumption are distributed.

If the models do not manage to result in realistic numbers and the analysis of data does not help to identify or resolve the problems (for example if variables have expressions far beyond the range represented in the samples and therefore cause extreme values), alternative solutions have to be considered. Alternative simulation-based approaches are explained in Appendix 13.17.

For both, crop processing and the empirical ME models, the aim is to conduct a sensitivity analysis at the end to test the robustness of the model and the sensitivity of the target variables to changes in the different parameters that have been selected for the model. A sensitivity analysis can help to respect the underlying uncertainties expected from the model and allow for an insight into the effect of differing assumptions. The sensitivity analysis could test the 5%- and 95%-confidence intervals of the covariate coefficients of the different models as well as a plausible variation of values of the independent variables.

4 Results for Crop Processing ME

A summary of the findings on electricity consumption are shown in Figure 7. A description of how to interpret these findings is given in this section, a more detailed description of the data, study selection and assumptions to compensate for missing data is given in the Appendix 13.3, as well as the exact values and a comparison with other literature on electricity consumption of crop processing (Table 14).

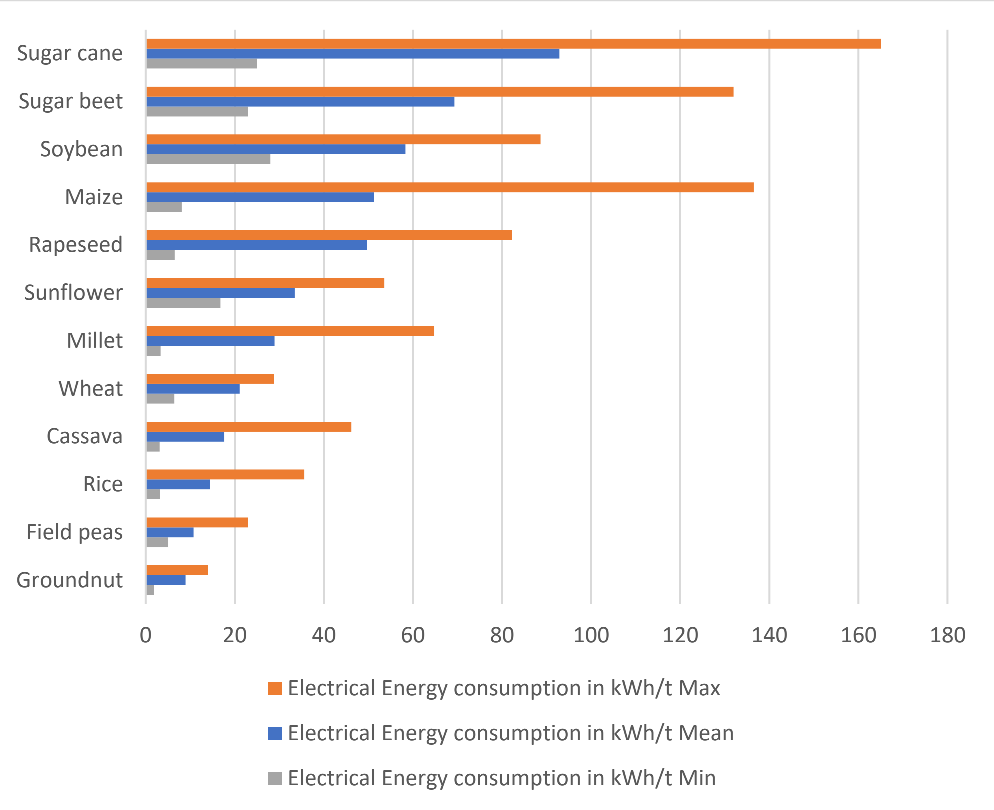

Figure 7. Average, minimum and maximum electricity consumption found per crop in kWh/t input.

First, it needs to be noted that only electricity consumption is considered, while the processes mostly also require other energy inputs such as thermal and manual energy. These other energy inputs were not considered since it was not possible to know in how far they could be reasonably substituted with electricity. Overall, the high discrepancy between the maximum and minimum values for electricity consumption stem from differing possible processing steps and differing possible electrification rates of the processing steps. For some crops (wheat, rice, millet, field peas, soybean), the consumption also refers to slightly different types of products, if they were relevant.

Despite these variations in processing operations and their electricity requirements found in the literature, the minimum, maximum and average values show mostly the same tendency for electricity intensity per crop. Producing sugar from sugar beet and sugar cane is a lengthy progress and requires by far most electricity. The need is in fact so high, that is unsure if it can be reasonably processed on a small scale as no literature was available on this scale. There are also less energy-intensive products that can be made from the crops (e.g. sugar cane juice) but there was not enough information given in literature to understand the relevance of these products or the electricity requirements of their associated processes. The high value for soybeans comes from oil processing, which is also rather elaborate, while the low value comes from a simple milling process. Maize was assumed to be processed to maize flour. Rapeseed and sunflower seeds were assumed to be processed into oil. Millet and wheat were assumed to be dehulled and milled to flour. Cassavas are usually purchased unprocessed but can also be transformed into starch at a relatively low electricity consumption. Rice was assumed to be either prepared to be milled to raw or parboiled rice. Field peas and groundnuts were just assumed to be dehulled, which requires little electricity.

5 Literature Review

The following chapters focus on the modelling of electricity demand for productive uses other than crop processing.

5.1 Table description and general findings

The literature review focused on identifying predictors of microenterprise uptake, performance and electricity consumption to set the frame for choosing data and modelling method. The final list of predictors is given in *Table 3*. Table 15 in Appendix 0 indicates respective literature sources per predictor so sources will only be explicitly quoted again in this section for additional background information. Both tables indicate whether a source referred to a variable in terms of its relevance to uptake (U), performance (P) or electricity consumption (C). In most cases, however, the predictors for all three dependent variables overlap and will therefore be considered jointly in this section. To get a better overview, the predictors were sorted into categories: household characteristics, individual characteristics of the entrepreneur or household head (HHH), enterprise characteristics, market access determinants (general, demand, supply and physical access), the general institutional framework, specific development program characteristics and miscellaneous. In reality, this differentiation is not as strict, e.g. the access to electricity can be a matter of household efforts but also depends on the local availability. In total, 61 predictors were found in 49 different literature sources. 7 predictors are HH characteristics, 10 are entrepreneur or household head characteristics, 5 are enterprise characteristics, 19 are market access determinants (6 for demand, 6 for supply and 7 physical access), and 16 concern governance and the general institutional framework. However, the predictors the last governance category were each generally only mentioned in single cases while many of the predictors in the other categories were mentioned more frequently. The most frequently mentioned predictors were education (16) and sex (9) of the entrepreneur, wealth of the HH or EP (13), access to finance (12) and the regulatory framework (6). Electricity supply was also a found as predictor for uptake and performance in many sources (11 times for access and 9 times for quality) but this could, to some, extent, also be due to the pronounced focus on this issue in the scope of this study. 32 of the predictors referred to business uptake, 42 referred to the performance and 23 referred to the consumption of EP.

Before diving deeper into the explanation of the different drivers, some explanatory notes about data quality and selection criteria should be made.

For all three dependent variables, uptake, performance and electricity consumption, it has repeatedly been pointed out in literature that drivers and predictors are difficult to determine due to the different contexts in which enterprises arise and the different needs for different types of enterprises, (Kooijnman-van Dijk, 2008; Rao, 2013; Renner, 2017). Hardly any of the predictors were ever found to have a consistent level of significance and strength of impact on either of the dependent variables but their impact rather depends on the given context. This becomes especially visible in studies which attempt to evaluate the isolated effect of one driver, e.g. the impact of electricity on business performance. Authors tend to conclude that findings are inconclusive because it is always a range of requirements which need to be met, as already explained in section 2.1. It is a major challenge of the present research to build a model in which the covariates are strongly enough related with the dependent variables to transcend the vast differences across SSA.

Moreover, it is possible that predictors interact by correlating compensating or enhancing each other, i.e. high-quality roads might compensate, to some extent, for a larger distance to the next urban centre. What is already known about the nature of these ambiguities is investigated in more detail in the course of this section while section 8 will explain how they were dealt with in the models.

It should further be mentioned that, in an effort to be as encompassing as possible, predictors of different levels of precision and measurement were included in the table. The table also includes predictors which have not yet been tested statistically. However, if a variable was tested and not found significant for at least one tested subset (e.g. country, area, township), it was not included. If a variable was only significant in some contexts, it was still included. Furthermore, some predictors are expected to be true drivers, i.e. have a direct causal impact on the dependent variable, some predictors might be approximations and representations of immeasurable drivers, some can be understood rather as correlated predictors than true causal drivers and relate to the depending variables e.g. as being side-effects, symptom of the same cause etc. Ownership of appliances such as a TV, for example, is more likely to be associated with uptake because both result from the availability of financial resources and decent market access rather than some causal effect TV ownership could have on EP uptake. The less straightforward explanation of relationships with these non-causal predictors was not considered as an exclusion criterion as they can still be powerful in explaining behaviors. Finally, many studies focus on obstacles and bottlenecks rather than drivers, but if the ideas were still useful, they were included in the table. These compromises were all accepted in order to be as encompassing as possible as it was anticipated that the availability and quality of data for the models would finally determine how predictors were used and quantified and probably narrow down the number of predictors significantly.

Table 3. Predictors for EP uptake, performance and electricity consumption as identified in literature.

Category	Predictor
Household Characteristics	Income/level of wealth/assets (U, C)
	number of hh members (U, P)
	number of rooms (U)
	shock experience (U)
	house size (U)
	agricultural activity (precipitation) (U, P)
	access to information (e.g. radio ownership) (P)
Entrepreneur Characteristics	knowledge/skills (general) (P, C)
	level of education (years of schooling, literacy) (P, U)
	digital and mechanical know-how (U, P, C)
	social networks (religion) (U, P)
	marital status (U)
	age (U, P, C)
	sex (U, P, C)
	awareness and acceptance of productive uses of electricity (C)
	personal character (P)
	level of entrepreneurial experience (P, C)
Enterprise Characteristics	wealth (income, assets, capital) (P, C)
	type/industry (P, C)
	size (P, C)
	registration and formality (P)
	location of enterprise (U, P)
Market Access (general)	
Market Access: Demand	population density (U, P, C)
	urban/rural migration (U)
	local wealth and purchasing power (P)

	local population/market size (P)
	number of people who transit (U, P, C)
	level of competition (P)
Market Access: Supply/Input factors	access to finance (presence of banks, interest rate, required collateral, required owners' equity contribution) (U, P, C)
	cost of input factors (other than electricity) (C)
	presence electricity access (U, P)
	electricity access capacity (voltage, AC/DC, …) quality (number and duration of blackouts in given time; presence of risk factors for supply) (U, P, C)
	cost of electrical connection (C)
	cost of electricity consumption (C)
Physical Market Access and Transportation	physical access to input goods (e.g. appliances) and services (e.g. maintenance) (P, C)
	level of urbanization (U, P)
	distance to urban centers (U)
	presence of roads and distance to roads (U, P, C)
	quality of roads (P, C)
	presence of regular market/commercial center (U, P)
	access to daily public transportation (U)
Governance, institutional context and development	conductive regulatory environment and policies (U, P, C)
	effort to start a business (U, P)
	level of corruption (U, P)
	political relevance and activity of locality (presence of communal administration) (P)
	level of social security (U)
	availability of subsidies to SMEs (U)
	Country (P)
	GDP (U, P)
	gender equality (U)
	presence of violent conflicts (P)
	access to water (P)
	presence of schools (P)
	presence of health facilities (P)
	Investment climate (U)
	access to cell phone communication (P)
Other	Season (U, P)
	weather (droughts) (C)
Development Program Characteristics	training and information provided and quality (BDS), knowledge sharing (U, P, C)
	monitoring progress and satisfaction (P, C)

5.2 Household characteristics

The level of wealth, income and capital of HH and EP have frequently been mentioned as crucial determinants for uptake, performance and electricity consumption. This is because access to external sources of finance is very limited and enterprises usually rely on their own savings or financial help from family and friends (ADA, 2016). Income is also the only household characteristic mentioned as predictor for electricity consumption of enterprises. Kooijnman-van Dijk found that entrepreneurs who already started off wealthy benefitted most from additional electricity consumption (Kooijnman-van Dijk, 2008). The identification of the number of rooms and the size of the house as predictors of enterprise uptake could also be associated with this issue as both tend to indicate more wealth (Nagler & Naudé, 2017). Another possible explanation for the predictive power of these last two variables is their association with a higher number of household members, which is generally associated with a higher rate of business uptake. This effect has been assigned to possible support from other members

(ADA, 2016) and the larger aggregated chance that someone might pick up a business simply due to the presence of more people. Furthermore, a larger number of HH members may allow to leverage more resources such as labour and finance (Alsos et al., 2013).

Two controversial variables are the experience of shocks and agricultural activity. The idea of the studies which report associations with agricultural activity is that there is either a negative relationship when a HH uses entrepreneurial activity to balance out losses from poor agricultural activity or a positive relationship when HH can use the gains from one activity to invest in the other, thereby growing profits from both (Owoo & Naudé, 2014). Self-employment has been shown to be used as means to balance out shocks which have an economic impact on a household. Conversely, depending on the business and the type of shock, such an event could also hamper entrepreneurial activity.

Nkegbe (2018) found radio ownership to be significant to explain sales levels which they assign to the potential of radios to provide relevant market information. They do, however, not provide qualitative evidence of this reasoning so the effect could also be reversed, i.e. higher sales allowing the acquisition of radios.

5.3 Entrepreneur characteristics

This section refers to characteristics of individuals that are associated with increased entrepreneurial activity and with characteristics of entrepreneurs associated with an EPs performance and tendency to consume electricity. The strongest predictors of an individual's propensity to engage in entrepreneurial activity as well as their performance and electricity consumption were the types and levels of knowledge and skills they possessed, such as literacy, formal education and entrepreneurial experience (Blodgett et al., 2016). Awareness and knowledge of productive applications and digital and mechanical knowledge were found to be important preconditions for electricity consumption but oftentimes lacking in newly connected areas (Mayer-Tasch et al., 2013). The ADA (2016) found personal characteristics such as working morals to be decisive for business growth.

The availability of social networks was reported to be important by providing general and financial support, information and improved market access, but were hardly ever attempted to be measured/quantified. Only Kooijnman-van Dijk (2008) found the type of religion to impact performance because of the social networks of some religion's communities in a qualitative case study in India. Perhaps for similar reasons, most studies found married individuals to be more involved in entrepreneurial activity than people without partner.

Other identified individual traits were not fully consistent. Generally, most studies found older entrepreneurs to be more successful, as proxy for experience (Efobi et al., 2019; Owoo & Naudé, 2014). However, Blodgett et al., (2016) report entrepreneurial activity after electrification to be higher amongst younger, more technically inclined entrepreneurs. Concerning gender, some studies report a higher share of male entrepreneurs (Akpan et al., 2013; Harsdorff & Bamanyaki, 2009; Mapako & Prasad, 2007) while others found a higher share of female entrepreneurs (Ackah, 2011; Canagarajah et al., 2001; Costa & Rijkers, 2011; Mead & Liedholm, 1998). A quite consistent finding was that male-lead enterprises performed better and had a higher chance to consume electricity. Lecoque & Wieman (2015) found that this was because more men were focused on opportunity-driven businesses and making profits while women were more involved in "survival business".

5.4 Enterprise characteristics

Enterprise characteristics can obviously only be true drivers for the productivity and electricity consumption. A rather consistent finding was that higher levels of wealth tend to further more accumulation of wealth and other performance indicators and larger businesses are also associated with better performance. Wealth has also been shown to play a role in determining the decision to get access to electricity and the level of electricity consumption. As will be discussed further in the next section, the cost of connection, consumption and appliances are a major barrier to use electricity for EP and HH alike. Formally registered businesses also tend to perform better as they usually have better access to resources according to Foster & Briceño-Garmendia (2010). They also found informal businesses to suffer higher losses in sales (20%) due to outages as compared to formal enterprises (5%), because they are less able to afford backup generators.

Another rather intuitive predictor for uptake, performance and electricity consumption is the type of business. There are some types of enterprises which are generally more present than others, perhaps because there are some sectors which are easier to enter than others and some types of businesses are more important than others[5]. Furthermore, some types of businesses, e.g. those which offer more elaborate/professional products or services, tend to perform better (Kooijnman-van Dijk, 2008; Mead & Liedholm, 1998). Equally, some types of businesses consume more electricity than others because their business operations rely to different degrees on elctricity. However, it must be kept in mind that not only does the type of enterprise determine performance and consumption but also the choice of business can be determined by the availability of electricity and other input factors which complicate or facilitate certain types of businesses. Some electricity-intensive types of businesses, such as welders, might only arise when they can be certain that the supply is sufficiently reliable. In this case, the expected level of electricity consumption can actually be seen to proceed the choice of enterprise type. As was pointed out in section 2.1, households which pick up an enterprise due to push factors tend to opt for easy-to-enter businesses, which do not require a large upfront investment and many skills but might also be less lucrative. In this case, businesses are already somewhat set up to keep a low profile when starting operation. Despite some complicating interactions in the causality between the factors, it can be said with some certainty, that the type of business is an important predictor for the different dependent variables.

A final predictor, the locality of operation, i.e. the type of building and if it is the owners' home, a marketplace, the street etc., is strongly associated with access to market, but on the decision-making level of the enterprise rather than with respect to infrastructure. It was argued In the PRODUSE report that businesses operated in commercial centres and markets or closer to main roads have access to a wider consumer base than those operating from home (Mayer-Tasch et al., 2013).

5.5 Market access general

Access to markets has been mentioned by virtually all investigated studies. Some refer more specifically to access to different market segments i.e. demand and supply factors, some to general physical access. The different determinants and forms of market access or lack thereof are explained in the following.

[5] Appendix 13.7.1 shows some distributions of different EP types. For example, small shops and other retail are consistently found to be dominating the business landscape across different regions.

5.6 Market access demand

In many studies on microenterprises in SSA, low demand has been named as main barrier to business operation, expansion and electricity consumption, even if supply conditions are unproblematic. For example, Kooijnman-van Dijk (2008) argues that electricity can only contribute if the EP has been at the limit of capacity by improving productivity, but usually there is simply not enough demand for additional goods and services.

An intuitive approximate measure for the volume of demand is the number of people as potential consumers implied by population density and the local market/population size. Associated with this are notions of migration, urbanization and the number of people who transit through the business location. Next to mere consumer volume, the Wilcox et al. (2015) emphasize that also the wealth and associated disposable income and purchasing power dictate the level of demand. The study also observes that low demand levels intensify competition between enterprises, which impedes business.

5.7 Market Access: Supply of input factors

Access to input factors is also a frequently mentioned constraint and is especially prevalent in more remote areas. In a broader sense, social networks, knowledge, skills and information can also be understood as input factors but the focus here in on rather external and material factors.

To some extent, it depends on the type of enterprise in how far the lack of different input factors is a constraint to operation. However, an omnipresent problem addressed in almost every single study is access to finance. The different associated barriers are absence of financial institutes in general, absence of credits and capital, poor functioning of existing financing systems and bureaucratic inefficiencies, unaffordable collateral and start-up capital, corruption, high interest rates and fear of EP of not being able to pay back instalments (Herrington & Kelley, 2013; Olawale & Garwe, 2010). New entrepreneurs mostly rely on internal funding of their own savings as well as those of friends and family members to set up their business. If the informal sector is factored in, this share is even higher (Bruhn et al., 2017). Despite the importance of this constraint, the only way in which the investigated quantitative studies addressed the impact of financial constraints was by factoring in wealth indicators of households and enterprises. Finding measures for the availability of financing options seems to be an issue.

Another form in which availability of finance impacts business operation is in the cost of input factors. Access to these factors is furthermore constrained by lack of local physical availability, or, in the case of input services (and some locally produced goods), lack of specific knowledge and skills due to absence of education and information. This has especially been addressed in studies on the consumption of electricity. Two major barriers to consume more electricity are high connection fees and the lack of affordable appliances, which could be overcome if appropriate financing schemes was available.

As the main focus of this research is the consumption of electricity, this issue has been investigated with special caution. Some studies name it as one of many infrastructural issues that can limit performance, but it differs from other types of infrastructure such as transportation in a sense that it can be used as a central input factor to some businesses. Furthermore, it cannot simply be used once an area has been provided with access. First, a connection needs to be established, which can be quite a cumbersome, time-intensive and expensive procedure. Furthermore, high costs of consumption and appliances can pose a barrier to use electricity and reap its potential benefits. Overall, evidence on the benefits for electricity on uptake and performance varies a lot across studies (about half of the studies find a relationship). This comes from the difficulty to measure the impact of energy as it is an intermediate input, with different degrees of centrality to business operations and limited directness of impact on business performance and thereby limited measurability. Again, the effect of electricity ultimately seems to depend strongly on the context. Relevant contexts given in literature are similar to the drivers that also promote electrical consumption in general. More precisely, this means that contexts such as the general economic activity of an area, type of business etc. can be expected to not only increase electricity consumption but also the extent to which higher levels of consumption lead to better business operation. E.g. Lenz et al. (2017) find impacts of electricity consumption in rural areas in Rwanda to be highest in areas which already had thriving business centers before access was provided. Several sources found different types of enterprises to vary in the degree to which they can be expected to consume and benefit from electricity (Lenz et al., 2017; Mayer-Tasch et al., 2013; Wilcox et al., 2015). The PRODUSE report, for example, found service EPs to connect soon after access had been established, use a wide range of appliances, especially in peri-urban areas and have high consumption levels compared to manufacturing EP, amongst which the connection rate remained rather low. The highest connection rates were found among electricity-reliant enterprises who started business after electricity became available and use it as core input asset, such as welding workshops. The PRODUSE report (2013) and Lenz et al. (2017) find an increased uptake of these electricity-reliant EP an increased diversification of businesses in electrified areas, which can unlock new demand markets and thus improve performance. Furthermore, in a qualitative case-study, Kirubi et al. (2009) finds potential spill-over effects between different sectors, e.g. agricultural EPs might only buy productivity-enhancing machinery such as tractors, if they know they have a workshop close by, which tends to be reliant on electricity. This is, however, hard to measure in a quantitative way.

Studies which could not find an impact of electricity on business operation (Muthoni, 2019; Wamukonya & Davis, 2001), assign this effect to weak uptake and only low-consumption usage (e.g. for light), which might not be enough to make the relatively high investments pay off. Grimm & Hartwig (2012) found that the supply of electricity could even be a financial burden, especially for marginalized rural economies. Harsdorff & Bamanyaki (2009) found an effect of electricity by SHS on uptake but not on performance and argued that the technology and according capacity used to provide electricity can also play a role for uptake. Some studies take a closer look into the actual quality of electricity access and find some evidence that the losses of an average firm (not just ME) in SSA to be 1.25 USD/kWh of interrupted electricity (equal to losses of about 80 hours per month) (Aithnard, 2014).

5.8 Market access: Infrastructure

The physical access to inputs and customer bases is associated with demand and supply but is so frequently mentioned and is expressed so diversely that it deserves its own section. The predictors can be roughly split for the availability transportation infrastructure and the local presence/distance to markets. Oftentimes, studies test for variables such as urbanisation level and distance to urban centres, which can be understood as an approximation for market access since urban centres are generally associated with higher economic activity, a wider customer base and a larger variety and reliability of input factors. A larger distance form urban centres, can, to some extent, be compensated by transportation infrastructure such as road quality, distance to roads and availability of public transportation.

Higher levels of urbanisation, closer proximity to urban centres and better transportation infrastructure have been found to generally have a positive impact on all three dependent variables, but there is some evidence that the relevance of physical market access differs for different types of businesses, as it seems to be less important for professional services, bars, restaurants and other services which serve mainly local customers (Christiaensen & Demery, 2017).

5.9 Governance

Studies further occasionally mentioned certain framework conditions provided by surrounding institutions (e.g. governmental or socio-cultural conventions in general) as well as the state of different general development aspects, which can impact business operation in more or less direct ways. The more general aspects given in literature, such as the regulatory environment, level of corruption, presence of violent conflicts, country and GDP can also influence the presence of other predictors such as transportation and electricity infrastructure. This can be useful if data on more direct indicators cannot be found but it can also cause some issues in the interpretation of variables due to explanatory overlap.

An example for a rather indirect indicator is given by Etim & Iwu (2019), who argue that if disadvantage of women was relieved, more entrepreneurial activity from their side could be expected. The presence of schools, water and health facilities are seen by Cabraal et al. (2005) as fundamental requirements for areas to be economically successful as these institutions ensure health and education and thereby the productivity of people as entrepreneurs but also consumers. Chu et al. (2007) state that the GDP is heavily dependent on the performance on microenterprises due to the importance of this sector in many developing countries. This means that even if this indicator is not a direct driver, it could serve as a predictor for performance. Social security is discussed in terms of crop insurance which can be a pull towards agricultural businesses while a lack of general good social security schemes can push HH into needs-based EPs. Similar circumstances can, of course, also hold for non-agricultural businesses. Concerning the investment climate, Aterido (2010) actually finds poor investment climates to add to the growth of micro-enterprises because these can serve as alternative to employment in larger enterprises whose relative growth is diminished as a consequence of the poor investment climate. This is another example of a push factor leading to business uptake.

The rather indirect influence of these predictors makes it quite uncertain in how far they can serve to predict uptake, performance and electricity consumption. Another issue is the measurement of these drivers as studies usually addressed them in more qualitative, conceptual terms rather than presenting measurable, quantifiable indicators. Olawale and Garwe (2010) for example, simply asked SMEs to rank perceived obstacles. It will therefore be explored in section 6 in how far the available data can serve to approximate and quantify these predictors.

5.10 Development Program Characteristics and Other Predictors

Many studies were conducted in the context of specific development programs whose success they were aiming to identify and therefore informed about relevant characteristics of these programs. This information is interesting to give some context to the other predictors but since it cannot be reasonably applied in the model, it will not be given further attention in this report.

Finally, some studies discussed the relevance of season and weather to some types of businesses, mostly agriculture-related businesses for which it only makes sense to operate seasonally, which, as previously discussed can also have implications for non-agricultural EP (Loening & Lane, 2007). Blodgett et al. (2016) further find electricity consumption to fluctuate over the different seasons of the year associated with weather and according consumption for heating/cooling . They also find droughts to have a significant impact on rural electrical consumption, which they assign to the interdependence with the agricultural sector. Depending on the composition of business types, it is possible that the intensity of different weather events will play a role in the consumption model.

There are some predictors for electricity consumption that have been left out of the model because they determine the supply rather than the demand side, e.g. risk factors to the grid and power stations such as violent conflicts which can lead to the destruction of infrastructure. These issues sure are relevant in determining the number of outages. There are however two reasons not to include them. Frist, the supply affects not only microenterprises but all consumers and should therefore be covered by the larger IMAGE model. It could be argued that this information could still be important to the model if the number of outages and losses in sales and electricity consumption due to outages had to be modelled with the purpose of assessing optimal demand, which is the target of the study (more information on the concept of optimal consumption can be found in in Appendix 13.13). However, it was refrained from this attempt due to the combined effect of quite highly abstracted relationships and the expected unavailability of data.

5.11 Conclusion

In conclusion, the literature review could provide some valuable insights into the mechanisms that drive the operation of microenterprises in developing countries. Despite some variables lacking quantification and measurability, the identified factors could already be used as a starting point to collect data and get an idea about the strength and direction of different predictors on uptake, performance and electricity consumption. Another main finding is that some interactions between variables can be anticipated which requires caution when interpreting coefficients. Furthermore, it is important to keep in mind the idea of push and pull factors for uptake and performance of EP as a significant source for possibly contradicting findings, whereby less favourable conditions can limit the presence of pull factors while enhancing the presence of pull factors. Lack of education, for example, can be positively related to uptake as a push factor because of limited employment alternatives, while the reversed case, a high education, can be positively related to uptake of opportunity-driven businesses because it can help to better access business opportunities.

The following section 6 will shed light on how it was attempted to find data to match the predictors while section 8 will show how the data was brought together in the regression models.

6 Data sources, transformations and quality

After having explored which variables could be important for modelling the consumption of electricity of a microenterprise, the present section is dedicated to explaining the different sources of data used, substantiate the choices for these specific data sets and explain how the data had to be transformed to be made useful in the models.

Table 18 in Appendix 13.5 shows which variables were found in which source for each of the predictors given in literature in their "raw" form. The table further contains some additional variables that did not fit in a specific category provided by literature but seemed to have potential to add some explanation to the different models, namely variables concerning the availability of digital infrastructure, presence of (alternative) employment opportunities, prevalence of poverty, economic/business activity and the energy intensity of an economy. In the table, some variables appear more than once because they can represent and approximate several predictors. For example, data on the distribution of wealth could explain both the wealth of a HH or EP or the purchasing power from consumers, possession of electrical appliances imply availability of necessary finance as well as physical availability, etc. This ambiguity of explicatory power of the respective variables will be explained closer in section 7 and the interpretation of the final models in section 8. Without double counting, 123 variables were found, whereby for some predictors several variables from different sources were found. The WB Household Survey provides 34 variables, the EP survey provides 9 variables, 25 variables could be taken from the DHS, 8 variables were taken from the WB Doing Business Studies, 26 variables are World Bank Development Indicators and the rest was taken from other sources. The important characteristics of these sources will be explained in this section. For 19 predictors found in literature (about 30%), no data could be found. However, some variables could not be tested in the regression models due a high number of missing values (the allowed share of cases for which a variable could have missing values was 10%) or lack of data for extrapolation for all of SSA, which is also indicated in the table. The final number of variables which could be tested in any of the models was 89.[6]

Some of the excluded variables are used in the descriptive statistics section in Appendix 13.9 to 13.12 to better understand the sample. The final table further explains the level for which the data is given, and the last columns indicate in which regression model which variable could be tested ("y" if testing was possible "n" if it was not possible due to data availability). In several cases, the "raw" variables had to be transformed to be useful. Those are listed in Table 19 with a small note about the type of transformation. For the most important transformations, more detail is provided in the according sub-section of the source. The table gives all variables that were finally used, transformed and not, from the same variable in the column "Model Variables". For original variables, the transformation type column was left empty. If the tested variables were used in the final models and/or the projection is indicated in separate tables per model in Appendix 13.14.

[6] For 26 predictors, no variables were tested. Some approximative explication could be delivered be the present variables, for example the share of female HHH could be related to gender equality and the number of broadband prescriptions could approximate access to cell phone communication but it was deemed too abstract to be further considered in the study.

6.1 Variables included in IMAGE

IMAGE provides projections for a variety of sociodemographic variables, which are time and scenario dependent and can be used as input data to feed into the final model (see section 2.3). These variables are the number of HH, the GDP PPP (Purchasing Power Parity) (2005 USD), population density (persons/km^2) and the HH electrification rate. These variables are assumed to vary according to the SSP2. Table 4 gives the trajectories of the variables in IMAGE until 2030 on the regional level.[7]

Table 4. Trajectories of time-dependent variables in IMAGE until 2030.

	Western & Central Africa	**Eastern Africa**	**Republic of South Africa**	**Rest of southern Africa**
	Population			
2015	471.65	293.53	52.80	154.09
2020	530.42	329.52	55.12	172.62
2030	655.76	402.66	58.93	211.06
	GDP_PPP			
2015	1839.81	1317.12	10213.56	2316.73
2020	2234.32	1619.98	12145.27	2782.92
2030	3400.23	2406.16	16407.20	3681.55
	Electricity access under baseline			
2015	0.47	0.29	0.80	0.34
2020	0.55	0.36	0.84	0.41
2030	0.69	0.50	0.89	0.54
	Average HH size			
2010	6.2	5	3.4	5.4
2015	5.9	4.7	3.2	5.4
2030	5	3.8	3	5

6.2 World Bank Household Survey

6.2.1 General description and purpose in this study

Over the past decades, the World Bank has conducted several surveys on the household level across many different countries covering a variety of topics such as general sociodemographic and socioeconomic information, labour, health, housing, education, food security, social security as well as general attitudes and experiences. To the knowledge of the author, there are no other similarly encompassing, detailed and representative databases that could have served the intended purpose. Countries in SSA for which such surveys have been conducted in the past ten years are Nigeria, Niger, Malawi, the Gambia, Ghana, Ethiopia, Uganda, Tanzania and Liberia. Finally, it was decided to only consider the 2018/2019 General Household Survey (GHS) of Nigeria[8], in the following referred to mostly as "HH Survey".

[7] Population density is given on the grid-cell level.

[8] The sample resulted from the fourth round of a larger panel survey a part of the WB Living Standards Measurement Study (LSMS) whose main goal is to improve agricultural statistics. For additional information on the surveys, see also he World Bank's website https://microdata.worldbank.org/index.php/catalog/3557/study-description.

The Nigerian GHS 2018/19 was chosen because it provides the widest variety of information, especially geospatial information, detailed information on household non-farm enterprises but also on other sources of income, some information on individuals living in the households as well as information on the community level.[9] The survey has also been conducted very recently and provides a large sample size (4979 households and 3919 enterprises). It was further seen as beneficial that the survey is part of a larger panel study which was considered to be used to build a time-series model. This idea was later discarded for reasons explained in Appendix 13.17. The main barrier to add information from the surveys for other countries was inconsistency of the study designs which would have led to a very heterogenous dataset with respect to included variables and sampling approaches.

To ensure representativeness on a national and zonal level, the sampling was conducted in a two-step random sampling method of enumeration areas and households within these areas except for some areas which were inaccessible due to security issues. To make the sample fully representatives, cross-sectional weights given for each household according to their probability of selection and adjusted for non-response ratios were calculated based on the inverse of their probability of being selected into the panel.

Due to the lack of cases without electrical connection in the WB EP survey (which is introduced in the next section) the data WEB HH survey data was used not only to model the uptake of enterprises but also for the propensity to have an electrical connection. This model was conducted on the EP level instead of the HH level. It was attempted to also use this sample to build the sales regression but the model from the EP survey had a better explicatory power.

6.2.2 Variables

Different types of income of HH were computed based on information from different parts of the survey (e.g. income from selling agricultural produce or employed labor) to better understand the sample (see 13.8). The numbers were given in Naira and have been calculated to equal USD 2020 (using exchange and inflation rates as given in

[9] When selecting the model, the idea was to try to model the decision of EP uptake on the different levels (i.e. the dependent variable being the share of HH with EP in one community). However, due to reasons specified in Appendix 13.17, it was only possible to model the uptake on the HH level.

Table 16). Several income variables have been computed: the total annual income, the income from the EP of all HH who have an EP (EP-HH) (i.e. HH which only have the EP as income source and HH who have income from an EP as well as other income) and the non-EP income of all HH (i.e. HH without EP and HH with EP and other income). All income variables have again been calculated per HH member. It is, however, not certain if all income sources were covered by the categories in the surveys.

Further information was given on the possession of different assets such as TVs, smart phones and regular mobile phones. This data will be used in the model for assets on which data is available for all of SSA and for those assets which can be understood as proxy for other predictors, i.e. a radio could indicate access to information.

There was further the indication if any of the HH members cultivated any crops, which will further be referred to as agricultural activity for simplification, but it does not include the cultivation of livestock.

Experiences of different types of shocks were provided on community level (environmental shocks, price shocks, etc.) and HH level (fire, theft, illness or death of HH member, etc.), but only those were selected for testing for which SSA-wide information was found, namely the experience of droughts and floods in the past three years.

Next to the variables on HH level, the survey also provides information on every individual household member as well as information what relationship each individual has to the household head. The EP survey also provides the individual reference number of the managers and owners of the EP. This allows to investigate some individual characteristics of the HHH for the uptake projection and of EP managers/owners for the connection projection (in 48% of EP, the HHH is the manager of the HH EP). Relevant individual variables are the sex, age, marital status, level of education and literacy. The marital status and education variables were given as categorical variables and were transformed to dummy variables to indicate whether or not an individual attended any kind of school or not and if an individual has any partner in the HH (married or not), so a clear dichotomous definition was given that could be used to compare with the DHS data.

EP-specific information used were the sales in Naira. Other than that, no EP-specific data such as age, registration status and ISIC code[10] could be used in the model as there is always a lack of data for the projection for all of SSA. They could, however, be used in a descriptive way to better understand the sample (see Appendices 13.8 and 13.9).

It was possible to obtain some preliminary georeferenced data for the location of the HH from the WB. For reason of anonymity, the coordinated were modified using a random offset of cluster centre-point coordinates or the average of coordinates of HH in an EA.[11]. The same data could be used on the EP level as the sample were filtered for those cases which indicated to operate mainly from home (43% of EP). Relevant georeferenced variables directly provided by the WB are the distances to the next major road, the next population centre and the next main market. There were also variables associated with agricultural performance such as annual mean temperature, annual precipitation, precipitation of the wettest month and quarter and slope. The HH surveys also indicated whether the sector of a HH

[10] short for International Standard Industrial Classification of All Economic Activities, a coding system by the UNO to standardize and structure industry branches according to Revision 4. For more information see their statistical paper: https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf

[11] The distance between the actual location and the indicated coordinate is 0-2km in urban areas and 0-5km in rural areas. For 1% of rural HH a 0-10km was applied to increase uncertainty while keeping noise low.

was rural or urban and in which of the 36 states or the capital city a HH was located. The information on location was also used to obtain data from further sources indicated below.

The sample proved suitable to explore the EP uptake and electrification rate amongst EP but since the surveys did not include information on electricity consumption, another database had to be found for this step.

6.3 World Bank Enterprise Surveys

6.3.1 General description and purpose in this study

Aiming to give a representative insight into the private sector of different economies to support policy making, the WB has collected data from 164000 firms of different sizes in 114 countries since 2005 in the form of their (more or less) standardized Enterprise Surveys (from here on referred to as EP Surveys). The questions investigate objective experiences and opinions on issues such as general firm characteristics, gender distribution, access to finance, annual sales, costs of inputs and labor, corruption, licensing, land and permits, taxation, informality, business-government relations, innovation and technology, and performance measures. Within the information on costs of inputs, EPs are also asked to name their cost for electricity.

The main focus of the surveys is on formal service and manufacturing enterprises with more than 5 employees in the cities/regions with major economic activity but there are also some country specific selection criteria, which made it possible to obtain data on microenterprises.[12] After filtering for studies on SSA with representation of micro-enterprises and studies which only focussed on one city, 14 surveys were left: Nigeria (2014; N=288), The Gambia (2006; N= 126), Namibia (2006; N=95), Swaziland (2006; 118), Uganda (2006; N=95), Burkina Faso (2009; N=87), Cameroon (2009; N=99), Cape Verde (2009; N=95), The Democratic Republic of the Congo (DRC) (2013; N=382), Rwanda (2011; 112), Mozambique (2018; N=105), Ivory Coast (2009; N=87), Kenya (2013; N=307) and Zimbabwe (2016; N=266). This gives a total of 2262 EP in the sample.

Despite taking into account some country characteristics, there are some severe limitations already inherent in the data, as the focus is on economically active areas, which excludes all EP in rural areas, which could also indicate a bias to wealthier regions and impair the representativeness of the sample. The sampling frames were usually a list of EPs, which were formally registered at national chambers of commerce. This focus suggests an that in many cases, informal firms are left out of the scope, which also biases the sample since these EP tend to face different challenges than formal firms and can also differ in other characteristics. As found in section 2.1, most microenterprises are rather invisibly operating in their homes so the mere requirement of visibility to be selected EP surveys means that the sample represents larger than the average EP. This problem will be further addressed in section 6.10.3. Moreover, the problems with obtaining high-quality sample frames also hindered the accurate calculation of sample weights. Stratifying the sample ensured that EPs of different sizes, industries and sometimes locations were covered in the surveys, but representativeness could not be ensured in most cases. Furthermore, the data is not georeferenced and not all surveys provide the name of an EPs

[12] For more information about the surveys, see also the World Bank's websites on the overall methodology https://www.enterprisesurveys.org/en/methodology and survey-specific implementation reports, manuals and questionnaires https://microdata.worldbank.org/index.php/catalog/enterprise_surveys

location so no further georeferenced variables could be tested. Another limitation was the inconsistency in questions and variable names which made the matching of information troublesome and, at times, even impossible. Some data transformations had to be done to retrieve the information of interest (see Table 19), of which the most relevant ones will be explained in this section.

The database was used to model the performance and consumption of EP. Since all EP in the model seemed to have a connection, it was not possible to also model the probability of getting an electrical connection and the HH Survey had to be used. Since no household data is provided the sample was not suitable for the uptake model either.

6.3.2 Variables

The only data from the EP Surveys that could be tested in the models were the size of locality, number of outages in a typical month, sales and cost of electricity.[13] Since the data was not georeferenced and the surveys did not include quantitative information about the different variables suggested to model market access, the size of locality was the only variable that could be used in the model. Latter was missing in the surveys from 2006 and was added manually when the city of operation was indicated. When only locality size and region were indicated, the city could be manually inferred in an attempt to make the location precise enough to use georeferenced data, but the resolution was still too coarse to be reasonably used. However, by knowing the approximate region of an EP, at least some more spatially refined data from other sources could be used.

To make use of the country variable in the model (and achieve applicability for all of SSA), they were summarized to regions: Central Africa (Cameroon, DRC), West Africa (Nigeria, The Gambia, Burkina Faso, Cape Verde, Ivory Coast), Southern Africa (Namibia, Swaziland, Mozambique, Zimbabwe), East Africa (Uganda, Rwanda, Kenya).

Some of the variables which could not be tested in the models were used to analyse and better understand the sample in the form of descriptive statistics (see Appendices 13.10 and 13.11). Furthermore, the sales and number of employees were used to filter EP according to the definition used in this study for microenterprises, i.e. a lower income than 100,000 USD/year and a maximum of nine employees, which left a sample of 2006 cases.[14]

Sales

[13] In some surveys, there are variables given which refer to the manager/owner of the EP, namely the presence of female owners, years of experience and the highest level of education. However, since female ownership and years of experience are no information that can be obtained for all of SSA and the level of education had too many missing values, these variables could not be used for the model. To account for the impact of gender parity and education, regional averages by the DHS were tested.

[14] Even though the definition of a microenterprise for the surveys was set to a maximum of 5 employees, due to problems with the sampling, the number of employees was higher for some EP and the sample and therefore had to be filtered.

The sales were given as total annual sales of the last financial year (FY) in the LCU (Local Currency Unit) The variable was transformed to equal USD 2020 to be comparable across the datasets.[15] Other studies which investigate performance focus on sales per unit of time or sales per worker in terms of labor productivity. However, in the EP survey, there was a high number of missing values for the hours of operation per week and the number of full-time employees (44% and 12%), no indication of the number of months of operation and only four samples had information about part-time employees. In modelling the total sales, it is implicitly assumed that the variables predict the overall size of a firm on terms of their hours of operation and number of workers. This limitation can be accepted since sales are just a means to calculate electricity consumption, it is not necessary to understand performance in this depth.

Electricity Consumption

Instead of asking for consumption in kWh directly, the surveys asked for cost of electricity in the last financial year and provided the values in LCU. Therefore, the variable was transformed to equal USD 2020 to be comparable across the dataset and then divided by the respective electricity tariffs in USD.[16] There are some problems with this approach, which might lead to a faulty representation of actual consumption and deteriorate the ability of predictors to explain differences. First, it was not always possible to find data for the exact year for every country so data for the next or previous year had to be taken. Furthermore, it was not known which tariff group an EP belongs to. Next, cost for connection as well as costs for other sources of electricity than the grid, such as fuel for self-generation (23% of EP indicated to have owned or shared a generator over the course of the last FY) were neglected. Mayer-Tasch et al. (2013) found that some people spend less for the same consumption as others because they share a connection with their neighbors and therefore only have to pay part of the connection cost. These considerations help to understand possible sources for residual scatter than the model cannot explain. If these arrangements are taking place in the sample, this might lead to some distortion in the estimated consumption as well. This effect should be rather limited because the cost per unit of consumption would still be the same.

6.4 World Bank Doing Business Indicators

Another source used by the World Bank were the Doing Business Indicators. For the enterprise survey, the Doing Business indicators for Starting a business were generally implemented on a national level but for Nigeria, data was also available on state-level. Included variables were:

---

[15] The transformation was done using the average exchange rate of the year prior to the year in which the data of the EP survey was published as approximation for the last financial year, which differed in definition for each EP) and the inflation rate of USD from that year to 2020. The sources and rates are given in

Table 16 in Appendix 13.5.

[16] The sources and rates are given in

Table 16 in Appendix 13.5. The exact dates which defined a financial year differed between firms, but only the tariff for one year was used, if possible, for the year prior to the year in which the data pf the EP survey was published. The calculation also involved the exchange rates to USD of that year and the inflation rate of USD of that year until 2020.

- The expected number of days it takes to start a business
- The expected number of procedures it takes to start a business
- The expected cost required to start a business (as a percentage of the economy's income per capita exclusive of value added tax (VAT))
- The expected capital required to start a business (as a percentage of the economy's income per capita exclusive of value added tax (VAT))
- The Distance to Frontier Score for starting a business, which indicates the gap between an economy's performance and the best practice
- The rank of ease of starting a business compared with other countries

Furthermore, the following indicators for getting electricity (again on national and state-level) were integrated to the enterprise survey:

- The overall score for getting electricity (simple average of scores for components except for price of electricity)
- The expected time it takes to get an electrical connection (calendar days)
- The expected number of procedures required to get an electrical connection
- The cost of getting an electrical connection (as a percentage of the economy's income per capita excl. VAT)
- The price for electricity consumption for an assumed consumption for a warehouse based in the largest business city for the month of January (in USD/kWh)

Some methodological decisions of the WB might limit the representativeness of the indicators for microenterprises e.g. the WB assumes a number of 10 to 50 employees and its location to be the largest business city of the country.[17]

6.5 World Bank Development Indicators

Some more general country-level developments indicators could be obtained from the Development Indicator Database of the World Bank.[18] Of course, those could only be reasonably implemented in the enterprise survey. I n most case, it was possible to retrieve data for the same years the respective enterprise survey referred to and the number of missing data was very low for most variables, which were retrieved. Some variables were only available for some countries and were therefore left out of the final models. No transformations were necessary for these variables

6.6 DHS

The DHS (Demographic and Health Surveys) Program of the USAID (the U.S. Agency for International Development) provides data on national and sometimes also subnational level of many countries on wide array of health, wealth and general socio-demographic issues. The DHS guide gives exact definitions of each variable so this section will only focus on the most important aspects. [19] Since the

[17] For more information on the methodologies used, see also the respective website of the World Bank. Starting a Business methodology.:https://www.doingbusiness.org/en/methodology/starting-a-business Getting Electricity methodology: https://www.doingbusiness.org/en/methodology/getting-electricity

[18] For more information on the methodologies and definition, see also the website of the World Development Indicators: https://datacatalog.worldbank.org/dataset/world-development-indicators

[19] For more information on the methodologies and definitions, see also the DHS website https://www.dhsprogram.com/What-We-Do/Methodology.cfm

surveys are not conducted annually, it was not possible to obtain the values for the exact same years which were investigated in the World Bank surveys. Nevertheless, a lot of valuable information could be obtained (see

Table 17 for countries and years). The same variables were extracted both for the household and the enterprise surveys. Some variables like the share of population with access to electricity could just be adopted without adjustments. The variables for which transformations were made are mentioned below.[20]

The wealth quintiles (WQ) are calculated by country by assessing a basket of goods and other properties of the household (e.g. number of people per room), which are assigned a weight and then compiled to a composite index per household. This index is then attributed to each person. People are then ranked and split into five groups (here WQ1 being the poorest wealth quintile, WQ5 the wealthiest). The data is provided as share of population per quintile on a state-level for each country and for the sum of all rural and urban areas. Unfortunately, there is no way to know to which WQ an EP or HH belongs because the DHS does not publish the exact definition of the WQ per country but only a vague description of the methodology. Furthermore, a categorical variable was calculated in both datasets to indicate which quintile is the most dominant in the state of residence, i.e. to which of the WQ do most HH belong. A problem with the WQ is that the exact boundaries of each WQ differ across countries so the WQ themselves only say something to the extent that for the same levels of wealth in different countries purchasing power parity is given. This limitation could, to some extent, be addressed, by including a balancing country-level indicator such as the GDP PPP, GNI PPP (Gross National Income) etc., or similar.

There are several variables given for men and women separately (e.g. education, household headship, ownership of mobile phone, bank account etc.) and some are only given for women. Where available, the variables were tested for the separate sexes as well as by an average value that was calculated manually. The variables for age were given as share of population by age groups of a ten year-steps. Based on this a variable was formed to represent the share of people between 30 and 49 years of age, which are the group most engaged in EP activity. Some data is given on the share of population who own a certain asset (e.g. bike, TV, refrigerator, etc.) which, just like the asset variable in the HH survey, can have different interpretations

6.7 Further Sources

Additional, georeferenced data was implemented in the HH Survey, for which the sources and descriptions are given in Table 5. Their purpose was mainly to specify the different forms by which physical market access is enabled or impaired, i.e. transportation infrastructure, distances to urban centers and local level of urbanization. Furthermore, nighttime lights were found to be a reliable proxy for electricity consumption in lower-middle income countries (Falchetta & Noussan, 2019).

Table 5. Further variables and their sources.

Nighttime lights	yearly average radiance ($\mu W \cdot sr^{-1} \cdot cm^{-2}$) recorded in the nighttime hours	(Elvidge et al., 2017)
Settlement type data	categories of settlement type: water grid cell, very low-density grid cell, low density rural grid cell, rural cluster grid cell, suburban or peri-urban grid cell, semi-dense urban cluster grid cell, dense urban cluster grid cell, urban center grid cell	(Florczyk et al., 2019)
Urban cells	number of urban cells in the 5-km radius around the coordinates	

[20] Unfortunately, no data was available for Cape Verde which had to be excluded in the modelling sample but was still considered for most descriptive statistics.

Distance to markets	distance in km to core grid cells	
Distance to the nearest urban center (>20.000 inhabitants)	In km	
Travel time to the next city (>50.000 inhabitants)	in minutes based on the fastest transportation mode, based on existing roads traffic etc.	(Weiss et al., 2018)
Road density	total length of roads in the 5-km radius around the coordinates	(CIESIN, 2013)
Distance to the next road	In km	
Occurrence of droughts and floods	In past three years	(Beguería & Vicente Serrano, 2016)

Moreover, the state-level human development index (HDI) was retrieved from the UNDP (United Nations Development Program) for both the household and the enterprise survey databases.[21]

6.8 Data Quality and Data Cleansing

After compiling the datasets, several quality criteria were tested to assess representativeness and validity.

6.8.1 Outliers

Since only secondary data was collected, there was already a diligent process of filtering out extreme cases and checking and correcting errors, which could bias the analyses, as indicated in the respective methodologies of each dataset. Therefore, few outliers were encountered. Since the sales were trimmed at 100,000 USD (of the respective year of survey), all potential cases with outliers on the right side were eliminated in both datasets. Since the sales are strongly skewed to the right side, no outliers were identified on the lower end. Electricity consumption also had most extreme values trimmed when excluding EP with sales above 100,000 USD. On top of this, three cases were removed which had a more than five times higher consumption than the next highest value (1,255,230 and 1,882,845 kWh/year as opposed to 205,779 kWh/year).

6.8.2 Missing values

For this research, a listwise exclusion of missing values was chosen in each regression as opposed to pairwise exclusion or imputation of values, i.e. if a case had a missing value in any of the variables in the model, it was excluded completely. This can lead to a relatively large number of excluded cases but allows to represent correlations amongst variables more accurately than the pairwise exclusion. Due to the evidence from literature, the occurrence of correlations seemed like a very likely issue and had to be addressed adequately. An imputation of values was tested for the EP sample but due to the

[21] The HDI is a composite indicator of life expectancy, education and the GNI PPP/capita. For more info on the methodology of the HDI, see the website of the UNDP http://hdr.undp.org/en/content/human-development-index-hdi

imputation of several inappropriate values, this attempt was discarded (see Appendix 13.6.1.1 for more details). The final size of the HH sample was therefore 4,781 and of the final EP sample 1,659.

The limit for missing values in this study was 10% of cases per variable, i.e. variables with more than 10% missing values were not tested in the regression models. In the HH sample, none of the relevant variables had more than 5% missing values and therefore no variables had to be excluded (see *Figure 34* in Appendix 13.6 for a summary of missing values for the final connection regression and *Figure 35* for a summary of missing values for the final uptake model).

However, some variables had to be excluded from the models of the EP sample. First, it was found that the DHS data had 10.4% missing values because it did not have any data for Cape Verde but due to the importance of the data, rather than excluding all DHS variables, it was instead decided to exclude Cape Verde from the regression models (leaving a sample size of 2006). This left a sample of 2,167 cases. The DHS data then had 6.5% missing values due to lack of data for two regions in Kenya and one unidentifiable region in Swaziland, which was deemed unproblematic. It was, however, found that the dependent variables, namely sales and electricity consumption had 13.5% and 35% of values missing (after excluding Cape Verde; see *Figure 36* in Appendix 13.6 for a summary of missing values in the EP dataset). At first glance it seems like the households with missing values for consumption simply did not have any consumption, however, it was found that most of the enterprises still indicated that they had faced some outages. Since the consumption was calculated based on the cost for electricity, it is likely that the households with missing values did not know or did not want to state their electricity cost rather than not having a connection. The same can be assumed for EP who did not indicate sales. It was investigated in how far there were differences in the independent variables between the subset of those EP with missing values and the subset without missing values. Little's MCAR Chi-Square test showed that there were some few but significant differences between the subsets. It was therefore attempted to impute values, but this did not lead to a large enough improvement of the regression model to justify the loss of transparency (see appendix 13.6.1.1 for more details). Therefore, the EP with missing values were excluded and the final sample was 1,609.

6.9 External validity and representativeness of the country choice for SSA

Efforts taken by the WB to make the samples as nationally representative as possible and some problems of representativeness have already been addressed but is also necessary to look at the representativeness of the sample countries for all of SSA in order to assess external validity. This means, it should be tested, n how far the data can be used to build a model that can be extrapolated and represent behavior patterns across the whole continent of SSA to work towards an accurate projection, especially because the limited data availability did not allow for a diligent selection process of countries and the sample had to be taken as given. The differences between countries are too large to undertake this comparison in a fully encompassing way. It can be tested, however, if the samples represent the full range of some selected conditions across the region as much as much as possible. Therefore, the chosen sample countries will be compared with the other SSA countries by means of some national-level development indicators, namely the GDP PPP/capita, HDI/IHDI, life expectancy and electrification rates.[22]

[22] It should be noted that the surveys were conducted several years in the past and rapid development changes might mean that the past values were in tendency a bit lower so actual values form the survey might represent a somewhat lower development level.

The country with the lowest GDP PPP/capita in 2019 according to the World Bank was Burundi with 782 current international $ (current int$) and the country with the highest was Equatorial Guinea with 37,400 current int$. Somalia might rank even lower than Burundi but there is no estimate available for recent years. Nigeria is ranking relatively high with a 5,900 current int$ while the EP survey countries span from 11,200 for Namibia to 800 current int$ for the DRC.

Measured by 2018 HDI scores, the lowest ranking country is Niger with 0.377, for the inequality adjusted HDI (IHDI) it is the Central African Republic with 0.222. The Seychelles have the highest HDI score of 0.801 and Mauritius has the highest IHDI score of 0.688. Nigeria has a HDI of 0.534 and IHDI of 0.349. In the EP surveys, Namibia has the highest HDI score of 0.625 and DRC has the highest IHDI score of 0.456. The lowest HDI score in the sample is 0.434 for Burkina Faso and the lowest IHDI score is 0.293 for Gambia (Human Development Report 2019).

With 53 years, the Central African Republic had the lowest life expectancy in 2018 while Mauritius has the highest with 74 years. In the sample, Cabo Verde has the highest life expectancy with 73 years and Nigeria has the lowest with 54 years.

As of 2018, Burundi had the lowest average rate of access to electricity of 11%, to which Burkina Faso is the country in the EP surveys closest to with 14.4%. In the Seychelles, 100% of people have access to electricity. In the sample, Cabo Verde ranks highest with a 93.6% access rate, followed by Eswatini with a 76.5% access rate. In Nigeria, 56.5% of people have access to electricity (The World Bank, 2018).

Overall, it can be said that the surveys neither cover the countries with the highest or lowest average wealth or development level but a far enough range to suggest enough representativeness with respect to the given indicators. Furthermore, the distribution of wealth within SSA countries also tends to be very unequal and poverty, wealth and development can take so many different forms that the given country sample might still have the capacity to represent a wide enough range of circumstances, given the sampling has been done representatively. Nigeria, for example, is one of the countries with the most inequal wealth distributions in the world with 69% of population living below the poverty line (Oxfam, 2017). If sampling was done to represent the whole range of wealth, then a vast spectrum of wealth and development can be expected to be covered. This comparison does not claim to present exhaustive evidence for the representativeness of the sample for SSA, as this virtually impossible to obtain due to the aforementioned heterogeneity of countries and forms of development. Rather, it should give a rough orientation for where the sample stands in comparison to the rest of SSA and explain the effort to make the sample representative as possible.

6.10 Representativeness of the Samples

In a second step, a closer look will be taken at the range of values of variables in the sample to check again, in how far representatives and external validity were achieved to understand in how far findings can be expected to be generalizable. Ideally, the distribution of values for each variable would be compared with the distributions in the populations, however, the frequency distributions for most variables in the population is not known. A potential lack of representativeness does not, however, have to impair the power of the sample to serve as an adequate sample to build a generalizable model (Rothman et al., 2013). Instead, it should be ascertained whether the categorical variables that will be applied in the model have all possible expressions and whether the range of values for scale variables span the whole range of possible or plausible values. For example, HH and EP from all WQ should be represented as well as different degrees of urbanization, levels of remoteness, levels of access to infrastructure, especially electricity, etc. Both male and female owned EP should be represented as well EP owned by people with different educational backgrounds. Where possible, distribution of data in the sample were compared to the population and with findings in literature to obtain a better idea of plausible values. The range of variables in the final models are given in *Table 22*, *Table 23* and *Table 24* in Appendix 13.7 and a general comparison of the values of the dependent variables with literature is given in Appendix 13.7.1. The main findings of this examination, especially those of concern, are presented in the following sections. If not otherwise specified, the range of variables were deemed fine. A more investigative analysis of the sample properties with respect to their predictive potential for the regression analyses is given in the descriptive statistics in the Appendices 13.8, 13.9, 13.10 and 13.11.[23]

6.10.1 Representativeness of the HH Survey

A central indicator in all models is wealth. It is therefore important to check whether the different levels of wealth are represented in the sample. This is done here by means of the DHS wealth quintiles WQ. It is important to remember that it is not possible to know to which WQ a HH or EP exactly belongs to but only the distribution of the area in which it resides. Ideally, the average distribution of WQ in the sample should be 20% per quintile but has been found to be slightly off in the HH survey (*Figure 8*).

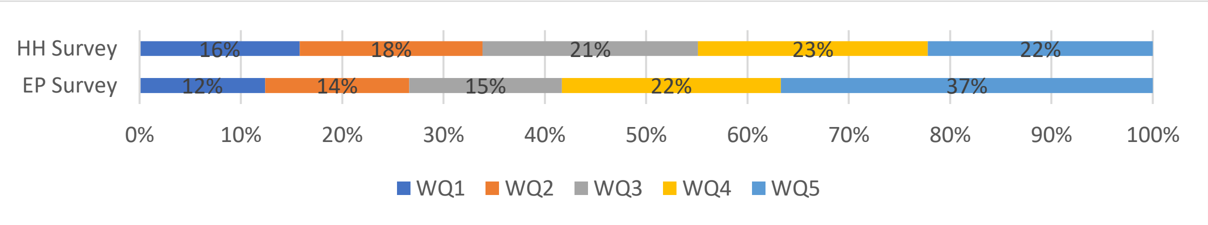

Figure 8. Distribution of WQ amongst HH in the HH survey.

Another dimension to look at is the distribution of WQ amongst in rural and urban areas. In the DHS data, there is a visible tendency in SSA for urban areas to be wealthier and rural areas to be poorer. This pattern is almost the same in Nigeria, which indicates a good representativeness of the country to the whole region (see *Figure 9* and *Figure 10*). In the HH sample, this tendency is much less pronounced compared to the whole population as suggested by the DHS data, but it still follows the overall pattern (see *Figure 11*).

[23] The Appendix will outline and analyze model-relevant structures of the samples, key effects of covariates on the different depended variables and relevant interactions between the covariates. This descriptive analysis helps to add transparency and a qualitative understanding of causal relationships as preparation before going into the regression analyses, where the effects of the "surviving" variables in the final models will be investigated. The analysis also addresses some variables that could not be implemented in the model e.g. due to data availability but are useful to understand the sample and possible confounding variables. In this context, descriptive analyses also helped to identify potential pitfalls for the analysis and limitations to develop coping strategies, e.g. for dealing with correlated variables.

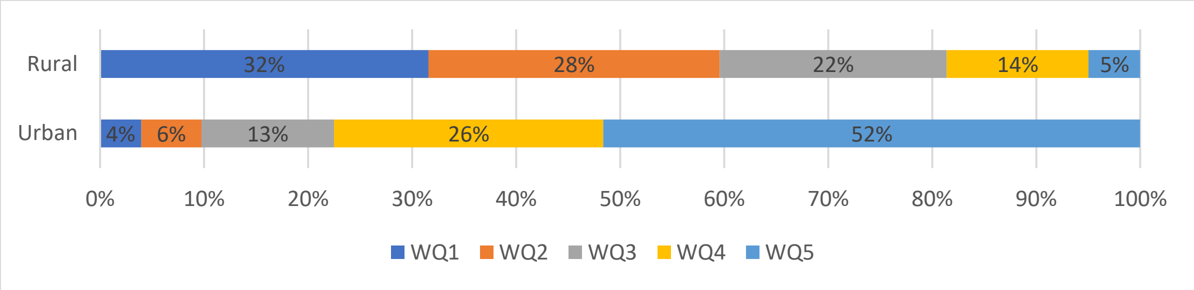

Figure 9. Distribution of population over WQ in rural and urban areas in SSA. Unweighted average of all SSA countries covered according to the DHS.

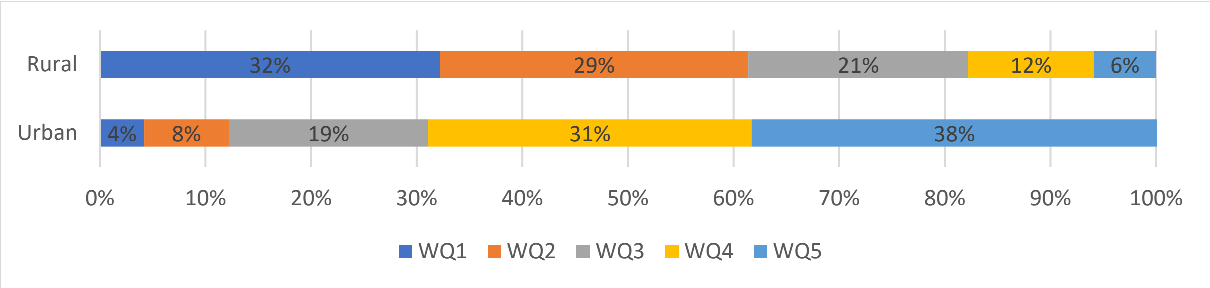

Figure 10. Distribution of population over WQ in rural and urban areas in Nigeria 2018 according to the DHS.

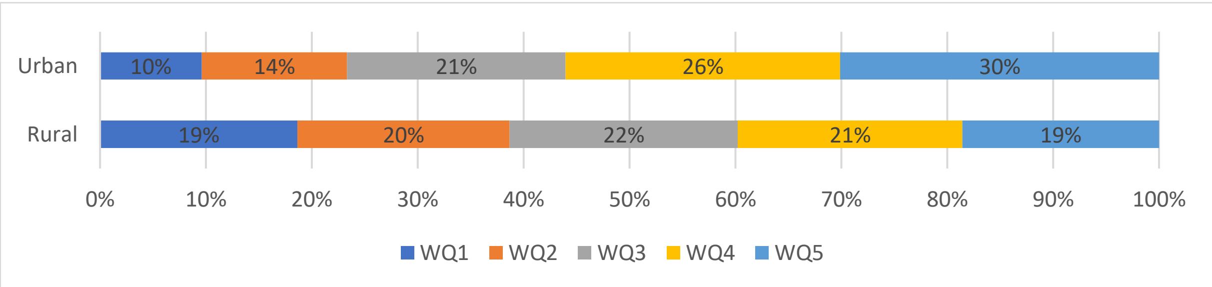

Figure 11. Distribution of sample over WQ in rural and urban areas in Nigeria according to the HH survey.

The share of population in rural areas in the sample (62%) also does not exactly equal the actual share of rural population in Nigeria (49%) (The World Bank, 2020), which indicates a weakness in representativeness. However, as previously pointed out, the regression function does not necessarily have to be impaired by this, as long as different degrees of urbanisation are sufficiently present in the model.

As electrification is of key importance in this research, it also makes sense to compare the sample electrification rate with the national rate. In the sample, 56% of households indicate having some sort of access to electricity. Of urban HH, this rate is at 83% and for rural at 37%. These values are representative for the electrification rates of 57% for the total population in Nigeria, 82% for urban and 31 for rural areas as given by the World Bank for 2018 (The World Bank, 2018).

6.10.2 Representativeness of the EP Surveys

There is quite an overrepresentation of the higher WQ in the EP surveys on an aggregated level, which is somewhat in line with the bias of the HH survey, but could lead to problems of underrepresentation of poorer EP. On the country level, it can be seen that this overrepresentation is taking place in most but not all countries.

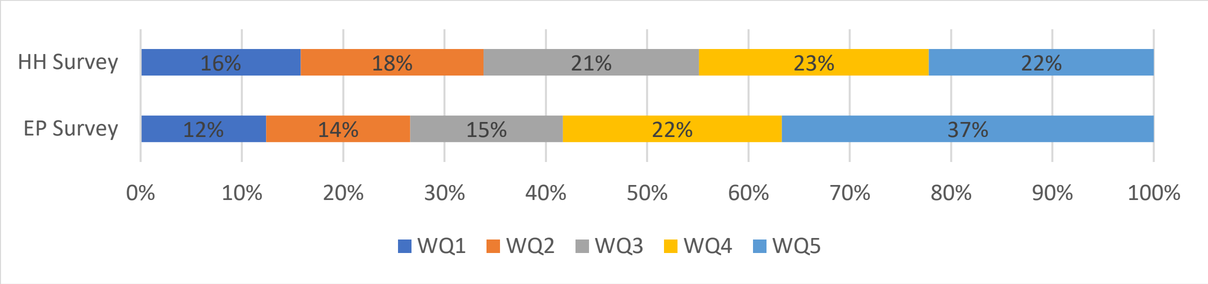

Figure 12. Comparison of distributions of samples over WQ between the HH Survey and the EP Survey.

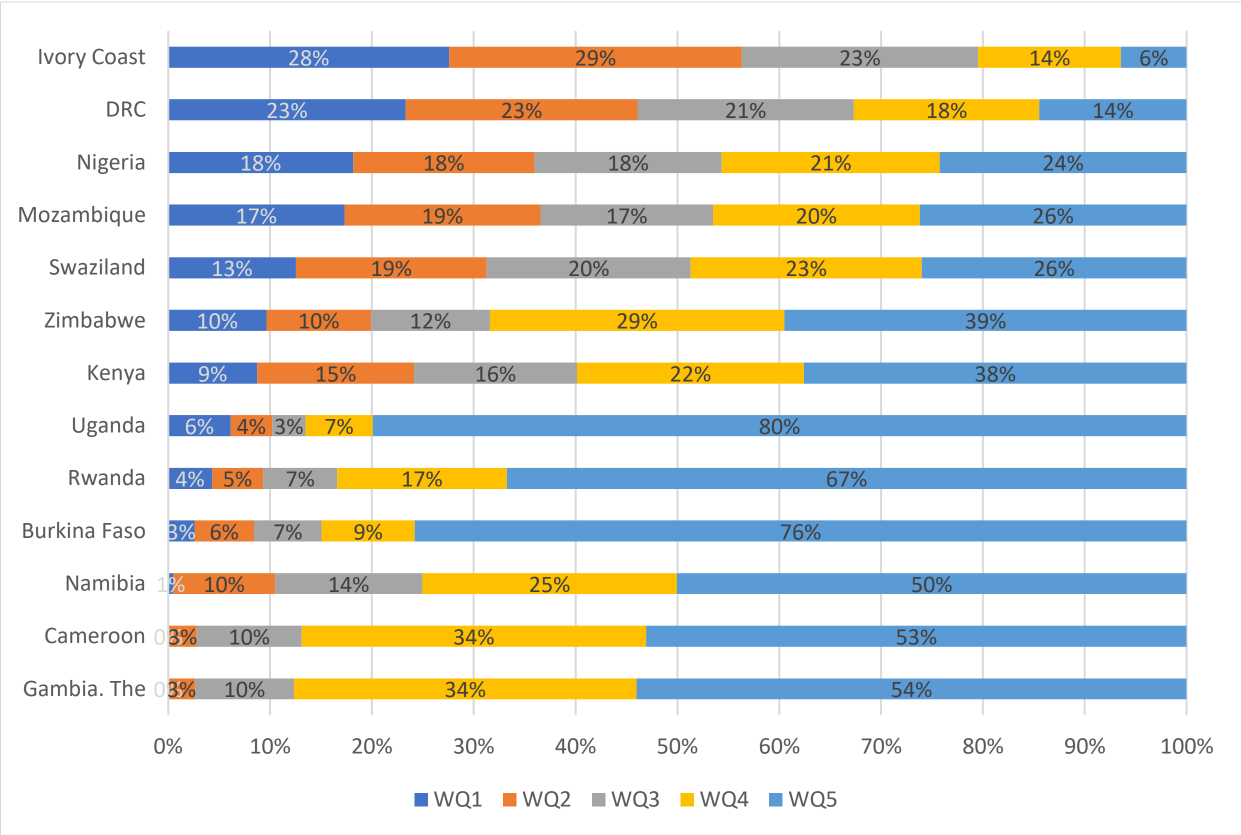

Figure 13. Distribution of WQ per country in the EP sample.

This bias can either mean that there is more entrepreneurial activity in the weather areas, or it might just be a result of an unrepresentative sampling strategy. Looking at the number of cases in the dominant WQ categories reveals that there is a severe underrepresentation of cases where the share of population in WQ4 is dominant, which can explain any unintuitive and inconclusive findings with respect to this wealth quintile. This biased distribution should be kept in mind for the interpretation of results but given that different levels are represented in the model, this should be enough to allow taking into account the effect of wealth on EP performance and consumption in the models.

Table 6. Distribution of EP sample amongst the different dominant wealth quintiles.

		Count
Dominant WQ in state of residence	1	392
	2	245
	3	220
	4	31
	5	1138

A differentiation of the indicators identified in the literature concerning the physical access to markets such as level of urbanization and population density could not be made because the database does not indicate the location per EP on a fine enough spatial scope. The variable on the size of locality does, however, indicate that the vast majority of EP were located in very urbanized areas (*Table 7*). While there is also a stronger tendency of EP to arise in urban areas in the HH survey, this very little number of EP located outside urban centers is clearly an underrepresentation of rural EP and a severe limitation which needs to be considered when projecting the data.

Table 7. Number of EP per locality size in the EP Sample.

	Count
capital city	348
over 1 mio	754
250.000 to 1 mio	831
50.000 to 250.000	253
less than 50.000	36

It was further found that most EP have an electrical connection. Counting all EP with either electricity cost and/or experience of outages gives a connection rate of 96% in the sample, which is higher than what has been found in literature.[24] The high connection rate has the advantage to allow a closer representation of consumption levels under universal access than if some areas/EP had no access to electricity but it also implies that some other conditions might be different between the sample EPs and the population, which can lead to a bias.

6.10.3 Comparability of the HH and EP Survey Samples

Another sample quality criterion to look at is the comparability of EP between the EP in the HH Survey and the EP in the EP Survey. Some remarkable differences were found, which might impact the modelling results. First, the average sales of EP in the HH survey (2,426 USD) are lower than those of the EP survey (25,087 USD) by a factor of ten. In principle, the drivers of sales can still be similar and modelled with the same variables so the different levels per se are not a problem, especially since the

[24] The mean actual connection rate calculate from available literature was found for be 56% and a potential connection rate to be 88%, i.e. the expected connection rate if there were no financial or other barriers and all EP which wanted to connect could connect. See appendix 13.7.1 for more information.

distributions of sales are very skewed to the right. The median sales of the EP survey are much lower (13,647 USD) than the mean, which means that a few EP with very high sales are responsible for the high mean while most EP have much lower sales. Furthermore, both surveys cover the range from almost 0 to 100,000 USD of sales per year. Nevertheless, it is a noteworthy difference, especially given the further findings. Concerning the number of employees, 11% of EP in the EP Survey employed one person, 20% employed two, 27% three, 30% four, 6% and the remaining 6% between 6 and 10. The HH Survey EPs are much smaller: 74% of EP only employ one person, 14% employ two, 7% three and the remaining 4% employ between four and ten people. In the EP survey, only 350 EP were asked more detailed questions about employment. Of those, 17% indicated employing at least one other member of the owners' family. Interestingly, also 17% of the EP in the HH Survey indicate employing at least one other HH member. Unfortunately, the time of operation is not comparable because the HH Survey only asked for months of operation while the EP Survey only asked for hours of operation in a typical week. It is furthermore noteworthy that almost all firms in the EP survey are formally registered (93%) and seem to have an electrical connection (96%) while in the in the HH Survey only 12% of EP are registered and only 62% have an electrical connection. Concerning the present business types, the same patterns prevail as in literature in both surveys (see *Table 25* in Appendix).

Some more comparisons are given in the form of descriptive statistics (Appendices 13.8 to 13.11) but it can already be said is that the differences in EP characteristics suggest that there might be a more fundamental/underlying, structural difference between the types of EP in the two surveys. From the mere samples, it is not clear, which of the "kinds" of EP is the more prevalent or can be expected to be more prevalent in the future. As they are both clearly still operating at a micro-scale, they are both interesting to the research. Due to the sampling method, it seems like the HH Survey displays the currently more frequent type of business that is largely operating informally, perhaps needs-driven on a very small, individual scale, while the EP Survey might display the a more "evolved" type of microenterprise which arise in more favorable settings (as suggested by the sampling method and the wealth distribution), are more formalized, have a better access to resources (such as electricity) and have a higher performance and a slightly larger scale (in terms of employees). In a sense, they could be seen as the version of microenterprise that can be hoped to be more prevalent in the future, if overall business conditions for microenterprises improve. Besides the uncertainty of relevance of the potentially different types of businesses, it is also uncertain, in how far their performance and consumption would be motivated by similar drivers. However, since this is the only available data, these uncertainties will have to be accepted and taken into account when interpreting the results.

6.11 Input Data for the Projections

For most variables which are not given in IMAGE, the same data sources are supposed to be used for the projection as the ones used for building the model. However, in some cases, there is a difference between modelling and projection data. If the model used categorical (dummy) variables on the HH-level, regional averages will be used for the projection. For example, the model might have been built on the indication whether a HH has an electrical connection (1) or not (0) (given in on the HH-level by the HH Survey) but data for the projection will be the average share of connected HH within the grid cell in percent (e.g. a value of 0.6 will be inserted if 60% of HH have an electrical connection instead of 1 or 0). Moreover, since IMAGE only has projections for some of the variables, the projections for the remainder of variables are just constant values based on current data. Furthermore, some approximations had to be made with respect to entrepreneur characteristics. The input data for literacy and mobile phone ownership has to be approximated with data on the population average, which, in reality, is probably higher amongst managers but it was still deemed a better solution than to leave out the variable. The same holds true for the ownership of mobile phones and agricultural activity, which indicate the share amongst EP in the sample while the data represents the share amongst HH. This has to be kept in mind for the evaluation of modelling results as a potential source of bias.

7 Electricity Consumption Modelling Theory and Operationalisation

Based on the findings in literature and own data analyses, the factors which influence EP uptake, performance and electricity consumption were conceptualized visually and will be explained below. The concepts outline the underlying assumptions made in the models are supposed to add some structure to the complexities and non-linearities between some the independent and dependent variables. The presented conceptualisations are the final versions and are therefore already adjusted according to the available data rather than to display the entire processes behind the determination of the dependent variables in the real situation. While some generalisations and approximations were made to capture the realistic causal processes, the concepts do not claim to be exhaustive. For example, the effects of some important but immeasurable, qualitative or very HH- or EP-specific factors (e.g. entrepreneurial experience, risk affinity, etc.) were not considered in detail.

7.1 Conceptual Model for Enterprise Uptake

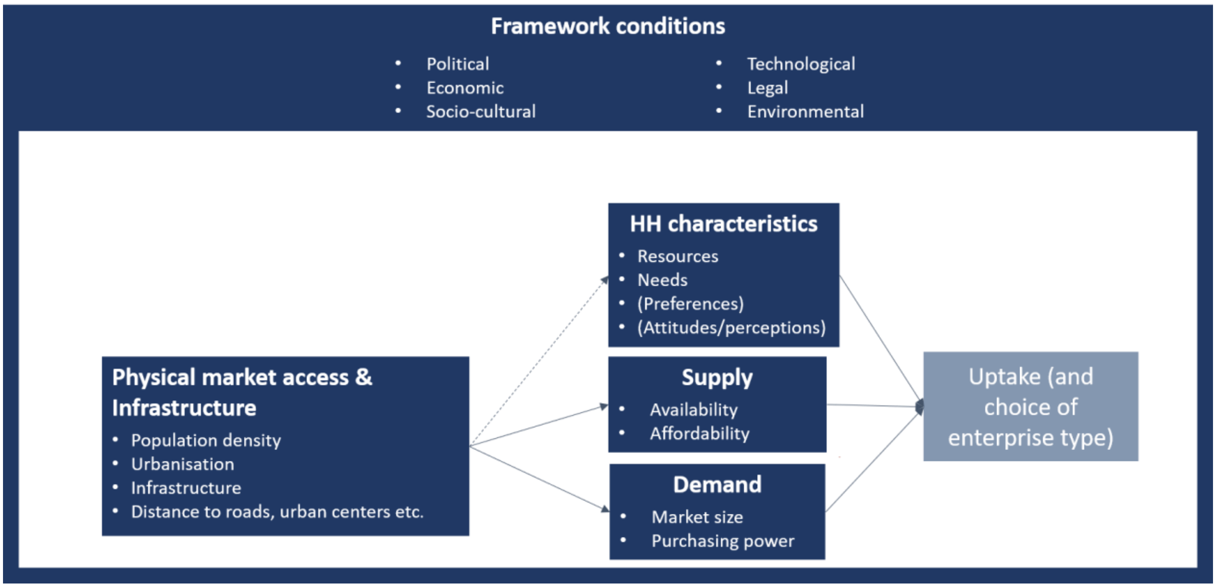

Figure 14. Uptake model conceptualization.

The uptake model concept captures the predictors presented in the literature review in their basic interactions and some additional relevant considerations. The presence of EP depends on the choice of HH to start a business and the external conditions, namely supply and demand (competition can be thought of as impacting demand), which determine business survival. For simplification, this model iss still referred to as “uptake” model. All three factors also depend to varying degrees on the physical conditions to access markets as well as other general framework conditions.

The framework conditions can be summarized as PESTEL criteria frequently used in conventional marketing research and strategic business planning They capture the different business-relevant external macro-environmental conditions, namely political (e.g. policies, political stability), economic (e.g. economic growth, interest rates), socio-cultural (e.g. age distribution, attitudes), technological (e.g. innovativeness), legal (e.g. consumer law, safety law) and environmental (e.g. weather, climate) factors. They can determine the uptake by influencing the different "channels" HH characteristics, physical market access, supply and demand.[25]

Due to lack of explicit data about the supply and demand of goods and services, they need to be approximated by the predictors provided in literature. For demand, this means taking into account variables associated with the market size (e.g. population density) and purchasing power (e.g. wealth distribution). For the supply side, this means taking into account predictors of availability (e.g. urbanisation) and affordability (e.g. cost) of input factors.

This need for approximation drives the choice of predictors more towards the macro-level (i.e. physical market access and framework conditions), which blurs out the explanatory power of these variables. For example, a higher population density might not only affect the survival of an EP due to a larger demand but can also be a predictor for a more reliable supply of input factors. The further the actual drivers need to be approximated and the level of abstraction increases, the harder it can be expected to find linear correlations between predictors and dependent variables due to other interactions. Some of the more macro-level determinants might even have reverse effects on the supply and demand. For example, a higher population density might also imply stronger competition and therefore decrease demand. However, due to the limited scope of the study and data availability, the concept and according regression assumes linearity between all predictors and the dependent variable. It needs to be acknowledged, however, that some predictors have rather complex effects. While the awareness of these interactions cannot find application in model itself, it can and must be considered for the interpretation of the results.

The decision-making process of HH depends mostly on internal HH characteristics, namely resources (e.g. social, financial, skills, etc.), needs, preferences, attitudes and perceptions. However, all of these factors can also depend on external conditions and the macro-level variables. For example, a higher population density might be associated with a larger number of schools which could improve the chances of HH to have the education required to successfully run a business. Even if perhaps to a lesser extent, the HH characteristics can also add to the above-mentioned difficulty to explain the effect of macro-indicators.

Further vagueness stems from the fact that, in reality, the decision-making process is made on the HH-level, but IMAGE works on a grid-cell level, which means that for each grid cell, the "average" household is modelled. The same conditions have to be assumed for all HH while in reality, for example a poor HH in an otherwise wealthy environment will probably behave differently than a poor HH in a poor environment.

Furthermore, the subjective nature of the decision-making process adds some complexity, which is rather implicit in the concept, as It was not possible to model it explicitly (therefore in brackets). EP uptake is ultimately a personal choice based on the subjective perception of the HH members about the importance they assign to different conditions as well as how favourable these conditions are. The predictors in the concept were chosen as rationally as possible with the aim to capture conditions which can objectively and generally be expected by HH to lead to successful business operation. But not separating the perceptions of HH from real conditions means that the suitability of the variables on external conditions to represent the decision-making process depends on the rationality and access to information of HH. In other words, it is assumed that actual favourability of business operation conditions would be perceived by the decision-making households as such and lead to activity accordingly, which might not always be accurate. Especially If a HH is pushed into business e.g. due to poverty, they might be less strict with their assessment of business opportunities.

[25] Possible feedback effects between the framework conditions and the market participants are not considered.

This leads to the general problem of ambiguity of certain predictors as addressed in section 8. The different reactions of push-/ pull-businesses to the same conditions are just one example, but it is perhaps even more accurate to say that generally, different conditions will favour the uptake of different types of businesses, be it in terms of motivations (i.e. survival vs. growth), their business operations or even location of operation. These differences in business types can have noteworthy effects on their performance and electricity consumption. An important question to answer would be by which characteristics EP should be differentiated to form categories that will best serve to explain how different conditions favour uptake for these different types of businesses as well as tendencies in electricity consumption. However, since the attempts to cluster business types did not work out (see Appendix 13.16.3 for more details), the present one-dimensional concept was finally chosen, hoping that the drivers of business uptake might be homogenous enough to still allow for building a representative model.

7.2 Conceptual Model for Enterprise Performance

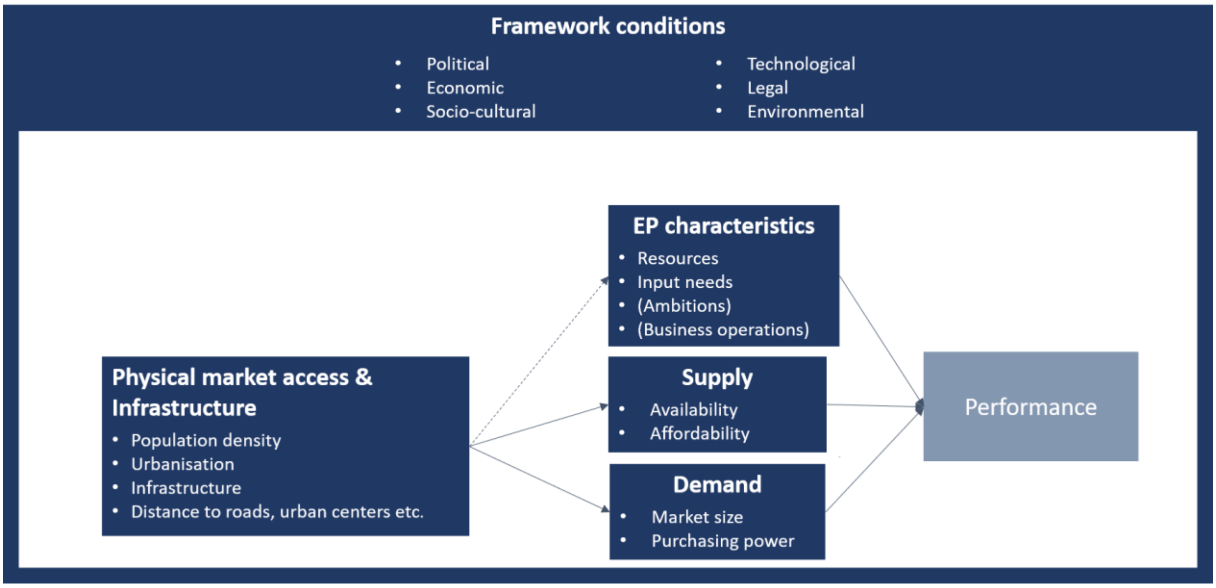

Figure 15. Sales model conceptualization.

The sales model looks similar to the uptake model as the drivers of success and prospective success/decision to start a business/survival are expected to be largely the same. Performance depends on the matching of the input needs of an EP and the supply of those inputs and the market demand for the goods and services provided by the EP. These might be quite different from business to business but again, due to lack of data, the modelling is done for an "average" EP rather than for specific business types. In general, the same limitations hold as for the uptake concept (explicatory power of macro-variables and non-differentiation of business types) except that in this case, the problem of potential discrepancy between subjective perspective and reality falls away since only the actual circumstances are considered rather than a decision process. To some extent, performance might also depend on the ambition of the entrepreneur but since this was not measurable for the model, it is kept in brackets in the concept, just like the business operations, i.e. type of business.

One input factor that is worth mentioning separately is the supply of electricity which is included as the number of outages in a typical month. By setting this value to 0 in the projection, the approximative performance under optimal supply conditions can be projected.

7.3 Conceptual Model for Electricity Consumption

Even though modelled in two separate regression analyses, the circumstance if there is a connection or not should largely depend on the same criteria as the volume of consumption itself as it represents the situation as if consumption is zero. Getting a connection is a more “extreme” case, since the conditions between having no connection and consuming the first Wh have to be marginally more favourable than the conditions between consuming the 1000th or 1001st Wh. Since the consumption model is set up in the two stages: the binary variable of having/not having access to electricity and the level of consumption, there is some room to account for this difference.

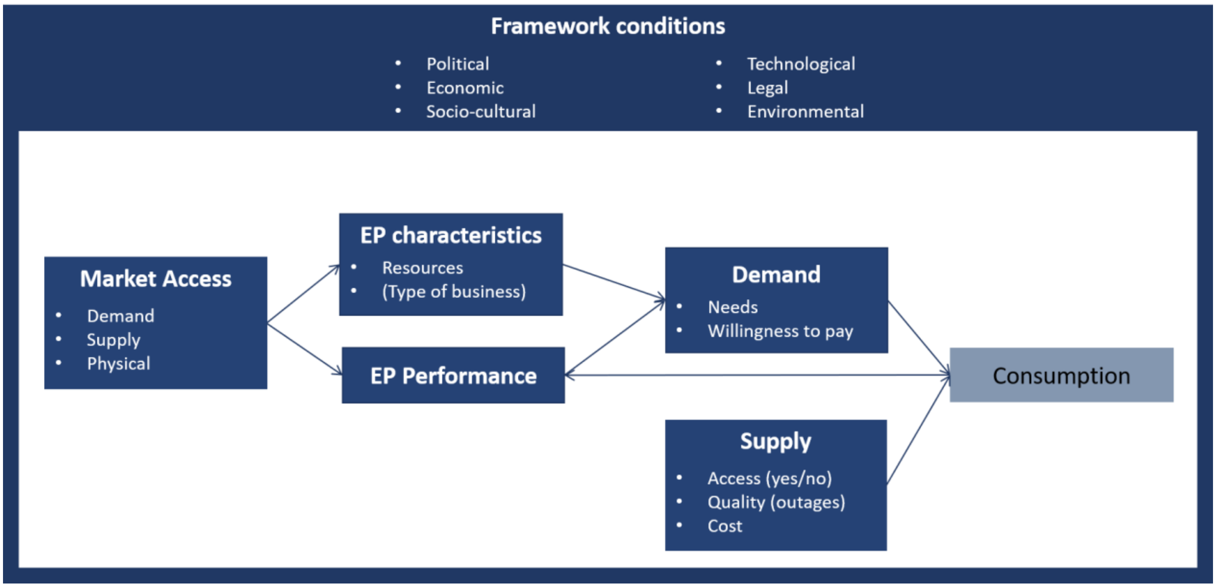

Figure 16. Consumption model conceptualization.

The concept for electricity looks a little more complicated than the other two but is simple in principle. Leaving out the factor of availability of appliances, the consumption of electricity, just as the consumption of any other good and service, is determined by the match of demand of the EP and the actual supply. Demand consists of the needed quantity and willingness to pay. The actual electricity supply provided consists of reliability/quality/quantity and cost of supply (i.e. the questions if an electrical connection is even possible, at what capacity, how reliable it is in terms of outages and how much it costs). The demand (volume and willingness to pay) depends on the type of business, i.e. how central electricity is as an input factor to operation, and the performance/size of the business which determines the required quantity of electricity. Both factors, as assessed in the previous two concepts, depend on market access (demand, supply and physical access are summarized here), and other framework conditions. Again, the type of business could not be addressed explicitly so an average consumption pattern has to be assumed. As an additional dimension, the balancing feedback between the consumption and the performance of EP has been added. This represents the impact that a mismatch of demand and supply has on the performance of the business. The activity of businesses has to be lowered to match the feasible consumption. This explains why a variable on the occurrence of outages is included in both models, sales and electricity consumption (see section 8). Since the aim is to model consumption under ideal conditions, the variables which indicate the number of outages in the two models can be adjusted to 0 respectively to model sales and electricity consumption under ideal conditions.

Just like for the previous two models, an overlap of some macro-level covariates can be expected which drive both supply and demand. While for the other concepts, this is a concern but not necessarily a barrier for the model to work, the consumption model should ultimately be able to project a scenario of perfect supply. If drivers of supply and demand are the same, it would therefore be important to know if (or to which extent) the effect of a predictor is taking place on the demand or the supply side, so the effect can be isolated to model optimal consumption. Concerning the consumption level, more information about the mismatch between supply and demand was attempted to be retrieved from some available variables that could have served to compute “symptom”-indicators of mismatch between demand and supply and compute optimal consumption. These mismatch-indicators were the interactions of vulnerability, electricity intensity, generator ownership and price elasticity, also with respect to different business types. Unfortunately, important data was missing to capture these concepts fully. Since the findings are still useful to understand the complexity of measuring the impact of outages accurately, the modelling attempt is explained in Appendix 13.13). To still have an idea which impact variables in the model have on the outages and ensure they are not too strong, Pearson correlations between the model variables and the number of outages were tested and can be found in Table 45 in Appendix 13.15.

8 Regression Models

After laying out the theoretical basis from literature (section 5), gathering according data (section 6), getting a qualitative understanding of the interactions within predictors and models (section 7), the present section will outline how the data was converged into the final regression models. For a first idea on the relevance of the variables, the linear relationship between the scale variables were tested using the Person correlation coefficient, the categorical variables were tested with the t-test for significant relationships with scale variables and the z-test for significant relationships with other categorical variables. The results are presented in Table 43, Table 44 and Table 45 in Appendix 13.15. However, due to multicollinearity and some potentially spurious correlations, not all significant correlations ended up in the final model. Indeed, different combinations of variables lead to different p-values and different directions (i.e. sings) and levels of impact in coefficients. After frequent test runs, only those variables were chosen which were both robust, (i.e. showed a consistently high significance and a consistent direction and magnitude of their coefficient in combination with different variables) and somehow logically explicable to avoid including variables with a random, spurious correlation (e.g. consumption per capita on the national level given by the World Bank was strongly negatively correlated with the electricity consumption of EP, which can be assumed to be a random correlation since not logically explainable). Another criterion was to avoid distortive interaction effects (for further explanations, see section 13.16.6). For some predictors, several variables were available which could not be implemented at the same time due to high correlation. In this case, the most robust ones were chosen and at similar robustness, the variable with higher significance was chosen. Table 39, Table 40, Table 41 and Table 42 list the tested variables with a brief indication on whether they were included in either the final model, only used as input data or for the projection in IMAGE or were completely excluded (and why). The following sections again wraps up the main key information about the regressions, explains the coefficients of the final models and explains how the models perform with respect to different quality criteria. A general criterion was to only accept variables at the 0.05 significance level. The acceptable level of multicollinearity was set at a Variance Inflation Factor (VIF) < 10, except for dummy variables which are inherently more highly correlated.

For all regressions, the cases were filtered for a maximum of 100,000 USD sales/year and a maximum of nine employees according to the formal definition of microenterprises in this research. Filtering for the criterion of a value of assets below 100,000 USD was not possible due to lack of data.

8.1 Uptake Logistic Regression Model

8.1.1 Interpretation of the model

The dependent variable in the uptake model is the dichotomous variable if a member of a HH is engaged in any entrepreneurial activity (=1) or not (=0). The value that is being calculated by the model is the natural logarithm of the odds that at least one of the household members will have at least one microenterprise. The first choice of independent variables is given in Table 39 based on the findings about predictors of enterprise uptake as well as some predictors which were originally mentioned to impact the other dependent variables but could reasonably be expected to have an impact on uptake as well. The final number of cases tested in the model was 4,781 unweighted and 25,965,581 weighted. The variables in the final model are given with their coefficients, p-values and exponentiated coefficient in Table 8.

Table 8. Variables in the final uptake regression function with their coefficients (B), level of significance and exponentiated coefficients (exp(B)).

		B	Sig.	Exp(B)
Constant		-1.029	0.000	0.357
HH characteristics	Number of HH members	0.061	0.000	1.063
	Share of pop. in lowest WQ in state	0.032	0.000	1.032
	Share of pop. in low-mid WQ in state	-0.014	0.000	0.987
	Share of pop. in middle WQ in state	0.002	0.000	1.002
	Share of pop. in mid-hi WQ in state	0.035	0.000	1.035
	Agricultural activity in HH	-0.156	0.000	0.856
	HH ownership of TVs	0.030	0.000	1.030
	HH ownership of mobile phone	0.332	0.000	1.394
	Experience of drought in community (3y)	-0.061	0.000	0.941
	Experience of flood in community (3y)	-0.032	0.000	0.969
HHH characteristics	Sex of HHH	0.110	0.000	1.116
	School attendance of HHH	0.509	0.000	1.664
Market access	Access to electricity in HH	0.485	0.000	1.624
	Travel time to the next city	-3.12E-04	0.000	1.000
	Sector	0.208	0.000	1.232
	Settlement type		0.000	
	water grid cell	-0.435	0.000	0.647
	very low-density rural grid cell	-0.172	0.000	0.842
	low density rural grid cell	-0.479	0.000	0.619
	rural cluster grid cell	-0.546	0.000	0.579
	suburban or peri-urban grid cell	-0.871	0.000	0.419
	Distance to nearest major road	0.009	0.000	1.009
	Distance to nearest population center	-2.11E-04	0.000	1.000
	Number of urban cells	0.009	0.000	1.009
	Distance to nearest market	-0.002	0.000	0.998
Framework conditions	Share of women employed in state	-0.026	0.000	0.974
	Share of men employed in state	0.016	0.000	1.016
	Doing Business: Expected days to start a business	-0.016	0.000	0.984

The non-linear nature of logistic regression models and the different scales of the variables make it difficult to directly interpret the coefficients (B). It can still be useful to look at the sign of the coefficient as well as the Exp(B), which is the odd ratio, i.e. the change in the total odds op EP activity by a change of one unit of the independent variable. For example, an increase in the number of household members by one unit (so one additional household member) increases the odds of a HH to have an EP by 0.063, holding all other variables at a fixed value. The intercept means that the probability of EP uptake if all variables were 0 is 0.357/(1+0.357)= 0.263, so 26.3%.[26]

It is possible to compare the dichotomous variables with respect to their effect because they have the same unit. School attendance has the strongest effect, changing the odds of EP uptake by 0.664, closely followed by the access to electricity. Agricultural activity decreases the odds by 0.244. The ownership of a mobile phone increases the odds again by 0.394. The ownership of a TV also still has a significant but much weaker effect. Experiencing droughts and floods has a significant negative impact on the EP even if not very strong. Finally, urban HH have a higher probability to start an EP than rural HH.

It is also possible to compare the impact of variables which are measured in percentage, i.e. the wealth quintiles and the share of employed men and women. They are all, however, around the same order of magnitude. The change of the share of one WQ implies a change in percentage of the other WQ so the variables are strongly correlated. Similar to the reference category of categorical dummy variables, one WQ (here WQ5) is therefore excluded. The increase of the share of population in the lowest WQ in an area by one percentage point increases the odds for uptake by 0.032. For WQ2, this change decreases the odds of uptake of an EP by 0.013, for WQ3 the odds decrease by 0.02 and for WQ4 by 0.035. This might not seem like a lot but considering the range of percentages is 100, the impact of some WQ is actually slightly disproportional, meaning the change by 1 percentage point might change the probability of uptake by a little more than one percentage point.

HH characteristics

As already indicated in the context of descriptive statistics, the coefficients indicate that the share of households engaged in entrepreneurial activity tends to be larger if the share of population in the lowest wealth quintile is larger and smaller in areas where more people belong in the highest quintiles. The relationship is not fully consistent across WQ but significant and could suggest that many enterprises in the sample are a result of push factors, rather than pull factors. The wealth could be interpreted as directly associated with the wealth of the HH but also as an indication for purchasing power of clients and the general economic development, which also impacts the supply side. The variable could therefore also be put under the category of market access. In this case, the lower purchasing power implicated by a larger WQ1 would be a contradiction to the finding in literature. It is possible, however, that the effect of a lower purchasing power is, indeed in place, and limits further business activity from push business "invisibly". The presence of both effects could also be a reason why the effect between wealth and EP activity is not linear.

The household size is positively associated with entrepreneurial activity, which is in line with the consensus on his variable effect in literature. The interpretation of the possession of appliances is less straightforward as they can have different meanings. They can be seen as a mere symptom of access to electricity but are not strongly correlated. They can also suggest the availability of appliances in general and a certain level of financial resources of a HH, which helps with business uptake. They also improve access to information. Especially the ownership of mobile phones suggests a certain level of digital know-how which can help with business operation. The negative sign of the agriculture activity

[26] However, this translation to probability only works with the intercept, as the change in probability by a fixed value for an odd ratio depends on the level of probability.

dummy for the household shows that the entrepreneurial activity is rather contradictory than complementary to agricultural activity, which was also ambiguous in literature.

HHH characteristics

The effect of education does not come as a surprise considering the evidence from literature and the descriptive statistics. It is still an important finding and could imply that even the more needs-driven businesses, which seem to be quite present in the survey, might require some educational background and could not be suited as an escape from poverty for everyone.

Female headed households are overall more likely to pick up a business in this sample while in literature, this was ambiguous, even though in specific circumstances, the reverse effect might be the case (see Appendix 13.8).

Market Access

The model supports the stream of literature which found a positive correlation between electrification and the uptake of businesses. Some caution is necessary though, as both, uptake and access to electricity might also be the results of other conditions, such as higher levels of wealth and development, rather than having a direct causal relationship. However, for the projection, it is already useful having assessed the mere predictive power, especially given that this relationship leads to a an exponential increase in consumption with increasing electrification, rather than a linear one (see Appendix 13.18 for a further explanation).

The uptake of enterprises is more likely to take place in urban areas. In combination with the observation that uptake is more likely in areas with a high share of households in the lowest quintile, this could hint that uptake is most likely in poor urban, perhaps slum areas. In line with this, the model also suggests that a higher the number of urban cells is associated with a higher likelihood that a household will start a business. It could be argued that this variable has too much overlap with the dummy variable on sector, but both are significant, have an acceptable VIF and add explanatory power to the model. It was therefore decided to keep both. Similarly, the distances to the next urban centre, major road and market were kept in the model even though they might, at first glance, risk to introduce too much multicollinearity but it was not the case, perhaps because they differ in nuances. They all suggest that closer distances favour the presence of businesses. As they can all be interpreted as representative indicators for market access, they confirm the relevance of this factor to enterprise uptake/survival as given in literature.

The negative relationship between the travel time and probability of enterprise uptake means that households in closer proximity to a city are more likely to start a business, which is also in line with better access to supply and demand markets.

Concerning the settlement type, being located in any of the regions in the model decreases the odds of entrepreneurial activity as compared to the reference category, the urban grid cell, which further confirms the previous findings.

Framework conditions

The number of days it takes to start a business is significantly correlated with enterprise uptake. The Starting a Business Rank was also significant, but it was only ranked on the country level comparing the 36 different states and the capital, thus not suited for the projection.

The occurrence of droughts is negatively related to the uptake of enterprises. It was first assumed in literature that natural hazards would be a risk factor to agriculture to which entrepreneurship was a response for balancing out this risk. Instead of being used to balance shocks, it seems like the harm experienced due to shock is more associated with limited business opportunities, perhaps restricted demand due to losses in agricultural income.

The share of women and men who are employed were entered as separate variables and interestingly, have contrary effects on the uptake of enterprises. While female employment is negatively associated, male employment has a positive coefficient. One explanation could be that this is due to simultaneity, i.e. one of the variables taking away some of the explanatory power of the other due to high correlation, but the variables turned out to be only weakly correlated. A more likely explanation could be that female employment tends to be associated with better education, equality and perhaps wealth, while male employment is more constant over different areas and circumstances, and rather associated with poverty.

After finding quite some ambiguity in literature around many factors due to the presence of push- vs. pull-businesses, it is interesting to see that favorability of conditions (such as education, wealth or the Doing Business indicators), which are only logically expected to favor pull-businesses uptake, have an overall positive correlation with business activity even though the examination of the sample provided quite some evidence for the predominance of needs- driven businesses. Conversely, it would also have been possible that unfavorable conditions meant that also the opportunities for other employment are so poor that enterprise uptake was chosen as a last-resort solution and therefore increased. It is still possible that these contradictory effects are taking place and weaken the effects of some variables, e.g. the DtF score for doing business was insignificant and the effect between wealth quintiles are not quite linear. This could be object for future examination, e.g. in a cluster analysis as suggested in Appendix 13.16.3). It is also possible that the data reflects the modalities which determine the survival of businesses to a stronger extent than it reflects uptake.

8.1.2 Quality criteria

Key quality data of regression models are the Nagelkerke R^2, which is a measure of the share of variance between the actual observed value of a case and the mean value of all cases that can be explained by the model, and the share of correctly classified variables. When building the model, the aim was to maximize these two indicators while also meeting the robustness criteria for the variables mentioned previously. Furthermore criteria (explained below) were tested for models in which the key quality criteria were met relatively well.

The Nagelkerke R^2 of the final model is 0.177, which is quite low but not uncommon in social sciences (Neter et al., 1990). The model managed to correctly classify 67% of cases, of which 80.8% of HH with EP were correctly classified but only 46.9% of non-EP HH were correctly classified which means the model has a tendency to overestimate the share of HH with EP. Unfortunately, it was not possible to achieve a better classification or Nagelkerke R^2. Overall, the expected share of HH with EP of 60% is the same as observed share. Nevertheless, the weaknesses of the model has to be kept in mind for the projection to explain potential illogical findings and perhaps adjust the modelling strategy.[27]

[27] A further look was taken at the Hosmer-Lemeshow test, a Chi-squared goodness of fit test, which the match of observed and expected event rates in population subgroups. The test is insignificant if the match is good, but the uptake model turned out significant. However, the test is discussed controversially in literature and criticized for measuring calibration error rather than goodness of fit and has therefore only been considered as additional information rather than a hard selection criterion for the model (Garcia-Valentin et al., 2015). Additionally, the Omnibus test, another chi-squared test, which tests for significant differences between the explained and

There are some requirements that need to be checked in order to perform a logistic regression correctly. One prerequisite is that the single observations need to be independent from each other, which can be assumed given the diligent sampling design of the World Bank.

Next, logistic regressions generally require little to no multicollinearity among the independent variables. This means that the independent variables should not be too highly correlated with each other. In order to check this condition, a linear regression was run to compute the VIF for each variable and ensure that the final model variables have values below 10.

A logistic regression also assumes linearity of independent variables and log odds. To account for this, the Box-Tidwell test was run and the correlations between the logits and the independent variables were tested for significance with the Pearson correlation coefficient. The Pearson correlations confirmed significant linear correlations. The Box-Tidwell test, for which interaction terms of the independent scale variables and their natural logarithm need to be tested in the logistic regression model, indicated some significant non-linear correlations (i.e. the interaction terms were significant). In trying to find some non-linear relationships to improve the model, transformations of the dependent variables were tested (e.g. by squaring them or taking their natural logarithm). As this did not lead to improved results (using the quality indicators mentioned in this section) the model was kept in its previous form.

Finally, logistic regressions typically require a large sample size. A general guideline is a minimum of ten cases with the least frequent outcome for each independent variable in the model. In this case, there are 29 independent variables, if counting the dummies of the settlement type each as one, and the lowest probability is 0.089, which means that a sample of (10*29)/0.089=3,258 cases should be given. Even counting the unweighted cases, this target is easily met.

unexplained variance, has been applied. This test confirmed that the model has a significantly (0.000) higher explanatory power than the null model.

8.2 Connection to Electricity Logit Regression

8.2.1 Interpretation of the model

The dependent variable in the connection model is the dichotomous variable if the HH of the manager has an electrical connection (=1) or not (=0). The value that is being calculated by the model is the natural logarithm of the odds that a HH or EP will have an electrical connection. The EPs were filtered for those who indicated operating their EP from home to increase the chance that the EP actually uses the available electricity of the HH productively. It is also possible that they use electricity if they operate outside of the home but since no information is provided, it is highly uncertain.[28] The first choice of independent variables is given in Table 40 and is based on the findings or drivers and predictors of enterprise uptake as well as some predictors which were originally mentioned to impact the other dependent variables but could potentially have an impact connection as well. The number of EP were filtered for sales below 100,000 USD and less than ten employees. The final number of cases tested in the model was 1,659, weighting was not possible (see section 6.2). The variables in the final model are given with their coefficients and p-values in Table 9.

Table 9. Variables in the final connection regression with their coefficients (B), level of significance and exponentiated coefficients (exp(B)).

		B	Sig.	Exp(B)
Constant		-2.951	0.000	0.052
HH characteristics	Agricultural activity in HH	-0.442	0.014	0.626
	HH ownership of mobile phone	0.494	0.000	1.689
	share of HH in hi WQ in state	0.043	0.000	1.049
Entrepreneur characteristics	Literacy of manager	0.608	0.000	1.896
Enterprise characteristics	ln Sales in the last FY (USD)	0.252	0.000	0.643
Market Access	Number of urban cells	0.043	0.000	1.043
Framework conditions	Experience of drought in community (3y)	-0.784	0.000	0.438

The intercept means that the probability of electrification if all variables were 0 is 0.231/(1+0.231)= 0.045, so 4.9%.

Comparing the dummy variables again shows that the experience of droughts has the strongest impact on electrification followed by the literacy of the manager, the ownership of a mobile phone in the HH and agricultural activity in the HH. Overall, the variables have stronger effects than in the uptake model. It seems like there is a clearer distinction between electrified HH and non-electrified HH than between EP HH and non-EP HH. More evidence is also provided in section 8.2.2.

HH characteristics

The effect of agriculture is negative, which would be explained by the associated lower levels of wealth amongst agricultural HH (see Appendix 13.9). It could also imply that the income from agriculture means that there is less needed to get an electric connection for productive uses. The ownership of

[28] The same model was also tested including those EP who did not operate from home (N=3773) to get an idea about possible differences but there were only small differences in the coefficients while the Nagelkerke R^2 and classification remained the same.

mobile phone ownership of the HH is positively associated with electrification. Again, there are many different explanations for this. Most likely, it is not a causal driver but rather an indication of a certain level of wealth and access to appliances, which also favours the decision to get an electrical connection. It could be argued that an electrical connection is a necessary condition to get a phone for charging, but the increasing number of charging stations in SSA implies that not all people with a mobile phone also need electricity in their homes. It is therefore also possible that phone ownership actually proceeds and motivates electrification and provides HH with information and the possibility for communication, which can unlock knowledge and resources for productive uses of electricity.

Again, the WQ variable was classified as HH characteristic as approximation for the wealth of an EP, but it might as well be an indication for the presence of demand and other favourable circumstances. Higher wealth of an EP increases their ability to afford an electrical connection. In general, a higher level of wealth in the location of residence of an EP can be an incentive for energy providers to establish a reliable supply, which encourages connectivity.

Entrepreneur characteristics

Confirming findings in literature, literacy of the manager is positively associated with electrification, possibly associated with higher wealth and better access to information about productive applications of electricity.

EP characteristics

For the sales, the natural logarithm was implemented as the values are skewed to the right and the linear relationship turned out more significant than for the regular value. An increase in sales is associated with a higher propensity to get an electrical connection, probably because poorer HH cannot afford the connection. Conversely, it could also mean that HH which have a connection can use the electricity to increase their sales via improving productivity/accessing new markets by offering new products and services, etc. Probably, it is a combination of both.

Market Access

Testing the different predictors for market access showed a relatively high correlation so only the variable with the strongest explicatory power was selected, in this case the number of urban cells, which is positively correlated with the access to electricity. The causality is ambiguous as it could mean both, a better supply, as well as higher demand for goods and services so that EP can actually benefit form improving their performance e.g. by substituting some labour-intensive manual work with electrified processes e.g. for manufacturing activities such as sewing. The supply of electrical appliances also tends to be better in urban areas, which could otherwise inhibit any ambitions to utilize electricity.

Framework conditions

The strong impact of the occurrence of droughts can also have different explanations. On the one hand, it could suggest a poorer supply of electricity due to lack of water to run hydro power stations ((Blodgett et al., 2016)). It was, however, found, that its impact on the electrification on EP, whose HH are also engaged in agricultural activity, was much stronger than on EP without any agricultural activity. While in areas without droughts, around half of EP with agricultural activity got an electrical connection, the occurrence of droughts reduced this share to 25%. The effect on non-agricultural household was much smaller, decreasing the share of 84% connected EP in non-drought areas by 8% to 77% where droughts occurred. As it was found that agricultural activity is associated with lower levels of wealth (Figure 17) is could mean that the presence of droughts further limits the financial resources of HH to afford an electrical connection by reducing their crop yields.[29] This can also limit demand for the goods and services EPs offer, which also limits their sales and ability to afford electricity.

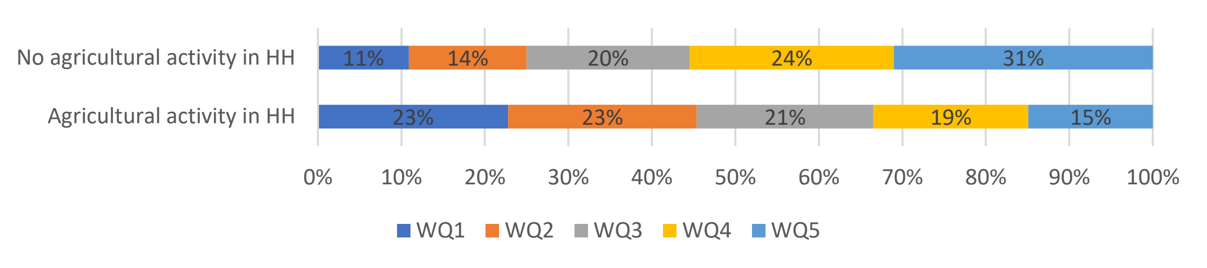

Figure 17. Differences in the distribution of wealth between agricultural and non-agricultural EP HH.

8.2.2 Quality criteria

As the connection model is also a logistic regression model, the same quality criteria and requirements were tested as in the uptake model. The Nagelkerke R^2 turned out quite high at 0.461, indicating a good explanation of variance of the regression function. The classification is also better than for the uptake model, overall classifying 77.5% of cases correctly of which 75.7% EP with electricity were classified correctly as having electricity and 79.6% of EP without access to electricity were classified correctly. Again, the overall expected share of EP with electricity of 55% is the same as the observed share.[30]

Concerning requirements for the correct application of the logistic regression analysis, independence of observations can again be assumed due to the careful sampling strategy of the World Bank. Too high degrees of multicollinearity were avoided by excluding variables with a VIF >10

To ensure linearity between log odds and the independent variables, again a Box-Tidwell test was performed and the correlations between the logits and the independent variables were tested for significance with the Pearson correlation coefficient Again, the Box-Tidwell test indicated that the relationships could also be interpreted as non-linear but the Pearson correlations confirmed the presence of sufficiently significant linear relationships (<0.05). No data transformations were found to improve the model.

The sample size is again unproblematic in this model as the number of variables is 6 and the lowest predicted probability is 0.066, which gives a minimum sample size of (10*6)/0.04=1,500.

[29] For more information on the presence and strategy to deal with interactions see section 13.16.6.

[30] The Hosmer-Lemeshow test turned out insignificant at 0.429, the Omnibus test was significant with 0.000.

8.3 Sales Linear Regression Model

The dependent variable is the value of annual sales indicated by the number given by enterprises for the total sales in the respective last financial years. Values were given in LCO and transformed to equal USD 2020. The frequency distribution of sales (see section 13.10) is strongly skewed to the right, which indicated that it might be better for the model to take the (natural) logarithm as dependent variable. More evidence on why the natural logarithm was chosen is given in section 8.3.2.

The first choice of independent variables is given in Table 39 based on the findings or drivers and predictors of enterprise uptake as well as some predictors which were originally mentioned to impact the other dependent variables but could potentially have an impact uptake as well. The final number of cases tested in the model was 1,609 unweighted, as weighting was not possible (see section 6.3). The variables in the final model are given with their coefficients, p-values and standardized coefficients in Table 10. The standardized coefficient allows for a better comparison of the relative importance of each of the variables in the model than the regular B. It corrects for different scales by calculating the coefficients as if the variables had been z standardized.

8.3.1 Interpretation of the model

Table 10. Variables in the final sales regression with their coefficients (B), level of significance and z-score standardized coefficients.

		B	Sig.	Standardized B
Constant		8.968	0.000	
EP characteristics	Power outages in firms in a typical month (number)	-0.049	0.000	-0.333
	share of pop. in lowest WQ in state	-0.011	0.022	-0.112
	share of pop. in low-mid WQ in state	-0.006	0.470	-0.051
	share of pop. in middle WQ in state	-0.003	0.619	-0.023
	share of pop. in mid-hi WQ in state	-0.033	0.000	-0.250
Market Access	National lending interest rate (%)	-0.099	0.000	-0.474
	Doing Business: Getting Electricity score	0.063	0.000	0.462
	Share of pop. without education	-0.008	0.000	-0.136
Framework conditions	SSA region - Central Africa	1.240	0.000	0.415
	SSA region -West Africa	0.759	0.000	0.251
	SSA region - Southern Africa	0.757	0.000	0.252

Comparing the standardized coefficients shows that the lending interest rate and score for getting electricity have the strongest predictive power. The SSA region dummies also have a high predictive power but varying noticeably between categories. The next important variable is the number of power outages faced by a firm in a typical month. The importance of wealth differs by quintile. Finally, the share of population without education also has a significant correlation with the level of sales.

The WQ can again stand either for the wealth of an EP itself or for the prosperity of clients or the activity and performance of the broader business environment. The impact of the WQ can be interpreted relative to the excluded category, here the highest WQ. Compared to higher shares of population in this WQ, higher shares in all the other WQ have a negative effect. This is a rather intuitive finding: a higher share of the lower WQ is associated with lower sales of an EP. However, as explained in Appendix 13.10, this effect is not linearly consistent over the different WQ. A higher share of the middle WQ, for example, would lead to a lower relative loss in sales as a higher share of the next highest WQ. A reason for this could be the unequal distribution of dominant WQ (see section 6.9) or the national-specific definitions of WQ, which impairs their comparability. For example, it is possible that someone in WQ1 in one country would be categorized in WQ3 in another country.

EP characteristics

The only significant variable on the EP level is the number of power outages faced by an EP in a typical month, because few specific variables were given in the surveys that had both, enough valid values (i.e. missing values) as well available data for the projection. The number of power outages is negatively correlated with the sales suggesting that businesses significantly rely on electricity access for their operations.

Market Access

The national average lending interest rate was introduced as an approximation for access to finance. A high interest rate suggests that credits are less affordable, which can inhibit business operation and performance. Given the strong emphasis in literature on the access to finance as a major obstacle for business operation, the high relevance of this variable is not surprising. An almost equally strong variable is the score for getting electricity as an approximation for the accessibility of electricity to enterprises in terms of cost and organizational burden. A higher score means that it is easier to get an electrical connection so the positive relationship with sales confirms the notion that businesses benefit from an easier access to electricity. This confirms the importance of electricity to business operation suggested by the outage variable. However, given the high level of aggregation and according abstractness of the score, it is also possible that, rather than representing the outlined causal effect, it could also just be a side-effect of a generally more advantageous business environment. Either way, its predicative power is undeniably given.

Framework conditions

The SSA dummy captures the effect that EPs in different countries have different levels of sales irrespective of other predictors, i.e. they capture some effects on the sales which cannot be explained by the other available variables. Another variable that was categorized as framework condition is the share of population without education. This variable implies that sales are lower in regions where people are less educated. While this seems intuitive, the actual effect is disguised. The most likely explanation is the absence of education is generally accompanied by lower levels of wealth on both the side of the enterprise as well as the customers, which decreases their purchasing power and effective demand. Lack of education could also be a hindrance to find other employment so areas with lower education have a higher share of low-income push-businesses. It could be argued that this variable represents market access or entrepreneur characteristics but because of the high level of abstraction and uncertainty about causality, it is classified as framework condition.

8.3.2 Quality Criteria

The key quality criteria the linear sales regression was optimized for was the adjusted R^2. A significant ANOVA (Analysis of Variance) (0.05 level) was used as knock-out criterion. The adjusted R^2 achieved in the final model is 0.26, the ANOVA was significant at 0.000.

To get a more practical idea about the predictive power of the model, the predicted and observed values can be compared. The average predicted value was 9.43, the minimum was 8.12 and the maximum was 10.788. Transforming the values back to USD 2020, the mean is 16,186, 3,578 for the minimum and 48,438 for the maximum value. This looks like the model has a tendency to underestimate the sales since the mean of the sample is 25,087. But taking a closer look reveals that the predicted values are just not as varied as the original values, which span from 124 to 130,638 USD, a common trait of regression functions referred to as "regression to the mean" (Barnett et al., 2004). The median values of the predicted and the observed values (12,100 and 13,646 USD) are quite similar, but the distribution of values is less skewed to the right. Indeed, while the standard deviation of observed sales is 27,596, This relatively small range of predicted values can be seen as a weakness of the model, but for the projection, this does not necessarily have to be a problem since it the aim is to look at more aggregated values on a grid-cell or even national or region level. At a higher aggregation level, extreme values tend to be crowded out anyway and the sales per grid cell are more likely to be closer to the mean than the observed values on the HH level. Moreover, the remaining underestimation might compensate for the suspected sampling bias addressed in section 6.10.3 and Appendix 13.10.

Next, conducting a reliable linear regression requires some conditions to be fulfilled, namely a linear relationship between dependent and independent variables, the normal distribution of residuals, homoskedasticity and the absence of multicollinearity.

The linearity of the relationship between the dependent variable and independent variable was tested using the Pearson correlation coefficient as given in Table 45. Similar to the previous models, the criteria for multicollinearity was the VIF. Again, the only variables with a VIF >=10 were the dummy variables, which are unproblematic.

The P-P Plots[31] in Figure 18 were made to check for the normal distribution of residuals. The graphs show that the natural logarithm of sales performed much better as dependent variable than the normal sales values and allowed to achieve a satisfying level of normality.

[31] The Ps stand for "probability", thus P-P Plots plot two cumulative distribution functions against each other, here the expected and observed probabilities.

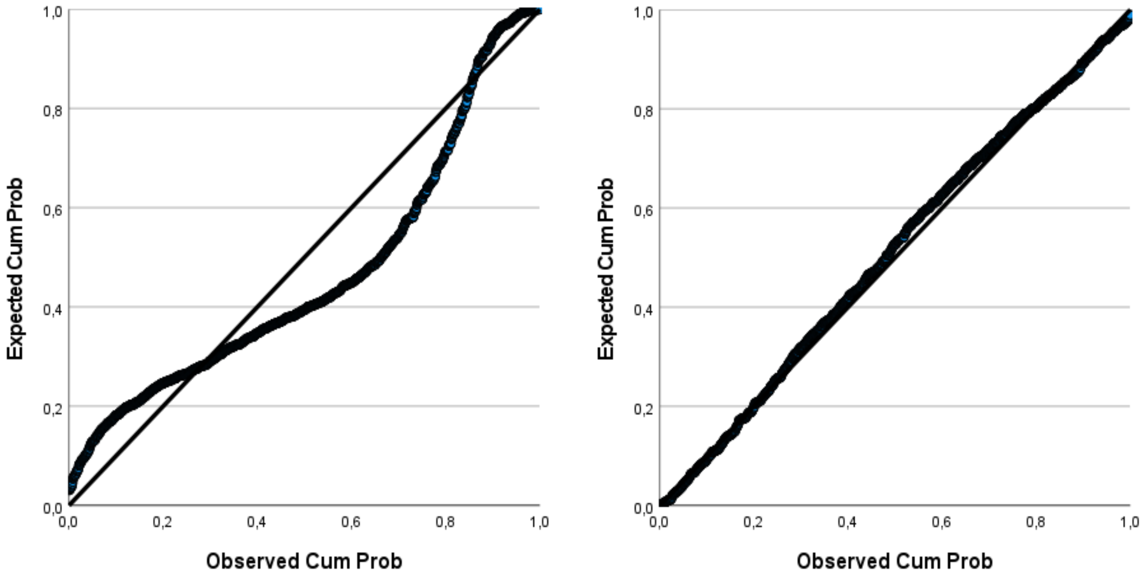

Figure 18. Normal P-P Plot of Regression Standardized Residual for the normal sales variable on the left and the natural logarithm of sales on the right.

The test for heteroskedasticity (the condition in which error terms are not equally distributed for different levels of the dependent variable) also confirmed the natural logarithm to be a better choice. While there is much more variation in the middle and mid-high range of values for the normal sales variable (see Figure 19), the values for the natural logarithm are more evenly distributed. Still, some heteroscedasticity could not be avoided. While for low and medium sales, there are about the same numbers of values below and above 0, there are more values below zero and more extreme values for higher sales, which means that the model is less accurate at these levels. This, in turn, can be either a symptom of impure heteroskedasticity, i.e. there are some effects, which are not captured by the model, or pure heteroskedasticity, which is the mere effect of more natural variability associated with the predictor variables. For example, a low education is usually associated with a low income as it is in many ways a barrier to the resources needed to gain a higher income, while people with high education can usually access higher income activities but do not necessarily do so. This leads to a larger variety of income levels for higher levels of education. A similar effect could be true for the relationship of the independent variables in the model and the level of sales. Since some important variables could not be added to the model due to data availability, the present heteroskedasticity is probably a mixed effect of both, pure and impure. This problem is not uncommon, especially in cross-sectional studies with a wide range of values of very heterogenous areas (Glejser, 1969) The problem of heteroskedasticity can be less precise coefficients but it does not bias them completely. In the context of this study, this can lead to somewhat less accurate projections. The two strategies to resolve heteroskedasticity are the addition of more variables and the transformation of data but since more data was not available and data transformations were not deemed adequate for the relatively low level of heteroskedasticity, the model was accepted in the presented form.[32] Since the major share of enterprises has a rather low sales level, the projection should still be relatively accurate most of the time. However, the present heteroscedasticity is clearly an issue that needs to be kept in mind.

[32] Transformations of the sales variable to reduce variability to e.g. labor productivity (i.e. sales/employee) or sales/hour of operation were not possible (see section). Other transformation options like weighting or the Box-Cox transformation were not deemed worth the potential bias this transformation would impose on the data.

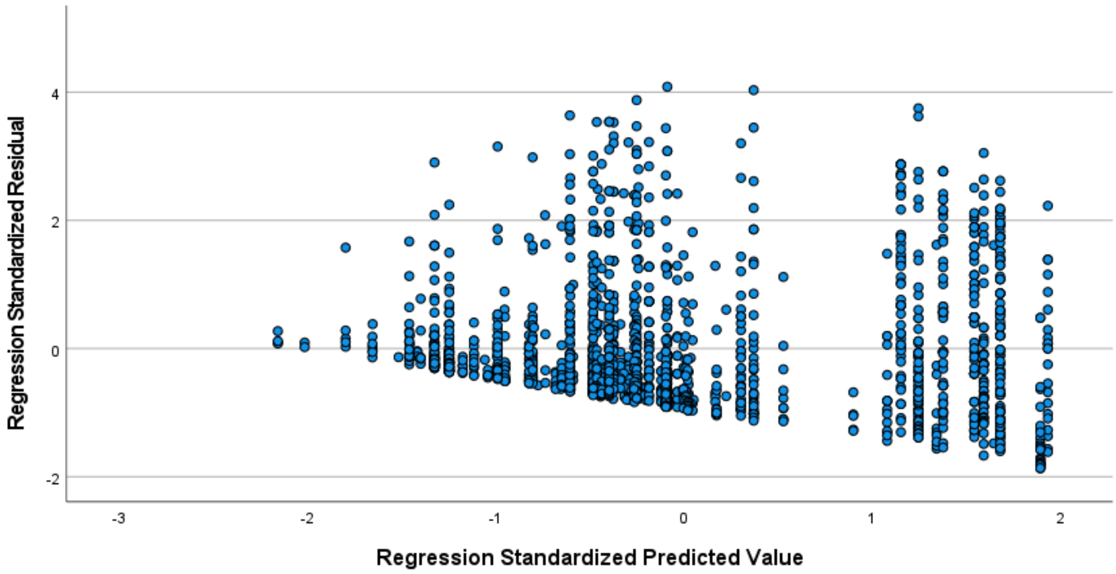

Figure 19. Scatterplot for the standardized residuals and the standardized predicted value calculated by the linear regression on sales.

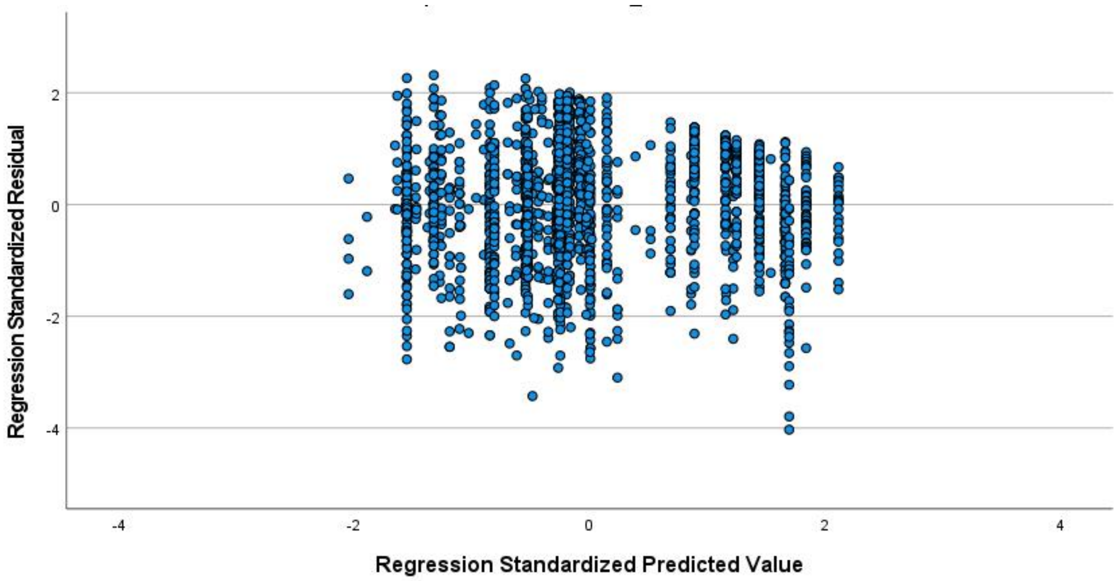

Figure 20. Scatterplot for the standardized residuals and the standardized predicted value calculated by the linear regression on the natural logarithm of sales.

8.4 Electricity Consumption Linear Regression

The dependent variable of the regression is the annual consumption of electricity of an enterprise in the last financial year in kWh. Following a similar reasoning as for the sales regression, it was expected that the natural logarithm would be better suited for the model as the distribution is strongly skewed to the right (see section 13.10). More evidence on why the natural logarithm was chosen is given in section 8.4.2.

The first choice of independent variables is given in Table 18 based on the findings of drivers and predictors of enterprise uptake as well as some predictors which were originally mentioned to impact the other dependent variables but could potentially have an impact uptake as well. The final number of cases tested in the model was 1,159 unweighted as weighting was not possible (see section 6.3). The variables in the final model are given with their coefficients in Table 11.

8.4.1 Interpretation of the model

Table 11. Variables in the final consumption regression with their coefficients (B), level of significance and z-score standardized coefficients.

		B	Sig.	Standardized B
Constant		2.837	0.000	
EP characteristics	ln Sales in last FY (USD)	0.430	0.000	0.288
	share of pop. in low-mid WQ in state	-0.058	0.000	-0.361
	share of pop. in middle WQ in state	0.022	0.025	0.106
	share of pop. in mid-hi WQ in state	-0.039	0.000	-0.202
	share of pop. in lowest WQ in state	0.020	0.014	0.145
Market Access	ln number of outages in a typical month	-0.173	0.002	-0.099
Framework conditions	Doing Business: Expected days required to start a business	-0.042	0.000	-0.637
	Doing Business: Getting Electricity score	0.047	0.000	0.244
	SSA region - Central Africa	0.488	0.029	0.108
	SSA region -West Africa	1.162	0.000	0.256
	SSA region - Southern Africa	0.420	0.092	0.102
	GNI, PPP (constant 2017 international $)	1.715E-12	0.000	0.151

Comparing the standardized coefficients shows that the variable with the most weight are the days for starting a business. Next comes the score for getting electricity and sales with about equal weights. The SSA region and WQ are also important but to different degrees depending on the category. Finally, the GNI and number of outages also impact the consumption significantly.

EP characteristics

As it was explored in section, the sales are positively correlated with the electricity consumption even though in literature the effect was not always conclusive (Mayer-Tasch et al., 2013). There are several possible explanations for this. Perhaps the additional conditions which need to be in place for EP to benefit from electricity for their operations (e.g. financial resources and access to appliances) were met to a larger extent in the present sample than in the (usually quite small) samples that were tested in literature. Another explanation could be that types of firms in the WB surveys had a more direct application of electricity as their core business, maybe electricity has been present for a longer time than in the cases observed in literature and is more established an input factor. Perhaps, the causality also lies the other way around and well-earning businesses are simply more being able to afford more electricity. A longitudinal study would be useful to investigate this relationship further. A similar reasoning can be applied to the WQ, which, again, need to be interpreted relative to the impact of the highest WQ. Their effect is not linear but significant. A possible reason for the blurred association could be due to the aforementioned issues that the WQ are not equally represented in the sample and definitions differ across countries.

Market Access

The number of outages (here the natural logarithm) decreases the consumption as was already discovered in section 0 by limiting the supply and keeping EP at consuming at their desired levels. If EP already know that supply is not reliable, they might also intentionally opt for business operations which rely less on electricity.

Framework conditions

The number of days to start a business and the score for getting electricity were found to be significantly correlated with the consumption of an EP. Other than having a direct causal link, it is possible that the days for starting a business stand for a generally more favourable business environment, which improves performance and thereby the size of business and extent of consumption. This assumption is supported by the fact that the score for starting a business was also significantly related with business uptake but had a too high multicollinearity. The suspected effect of the score for getting electricity is self-explanatory: A lower cost and administrative burden favour consumption. The GNI PPP is also positively correlated with consumption, i.e. EP countries with larger economies tend to consume more electricity. This is probably also associated with higher wealth. It does, however, not correlate with the WQ as these are country specific. It is possible that it adds some context of national wealth differences, which are ignored by the WQ. The SSA regions capture further country and region-specific differences that could not be explained by other variables.

8.4.2 Quality Criteria

The adjusted R^2 of the final model is quite high at 0.39, the ANOVA is significant at 0.000.

The mean predicted value was 6.40, the minimum was 3.35 and the maximum was 10.39. Transforming the values back to kWh, the mean is 1,656, the median is 851, the minimum is 28 and the maximum is 32,490. The observed mean is 5,681, the median is 1,712 the minimum is 18 and the maximum is 205,779. The model has a clear tendency to underestimate the consumption. Again, it also provides a narrower range of values with a standard deviation of 2,590 as compared to the observed standard deviation of 12,487. However, as already explained in section 8.3.2, the latter issues are not necessarily a problem for the projection. The bias to the low end, however, might be a problem in the projection and alternative solutions might need to be considered (see section 13.17).

As for sales, the basic requirements for conducting a linear regression analysis were checked. The linear relationship between predictors and dependent variable was tested prior to the implementation of variables to the model in the form of the Pearson correlation coefficients given in Table 45. As for the sales model, the multicollinearity was checked by means of the VIF for cases ≥ 10 and again only the dummy variables exceeded this threshold.

To check for the normal distribution of residuals, P-P Plots between the expected cumulated probability and the observed cumulated probability were made. As expected, running the regression with the natural logarithm of electricity consumption lead to much better results as can be seen in Figure 21.

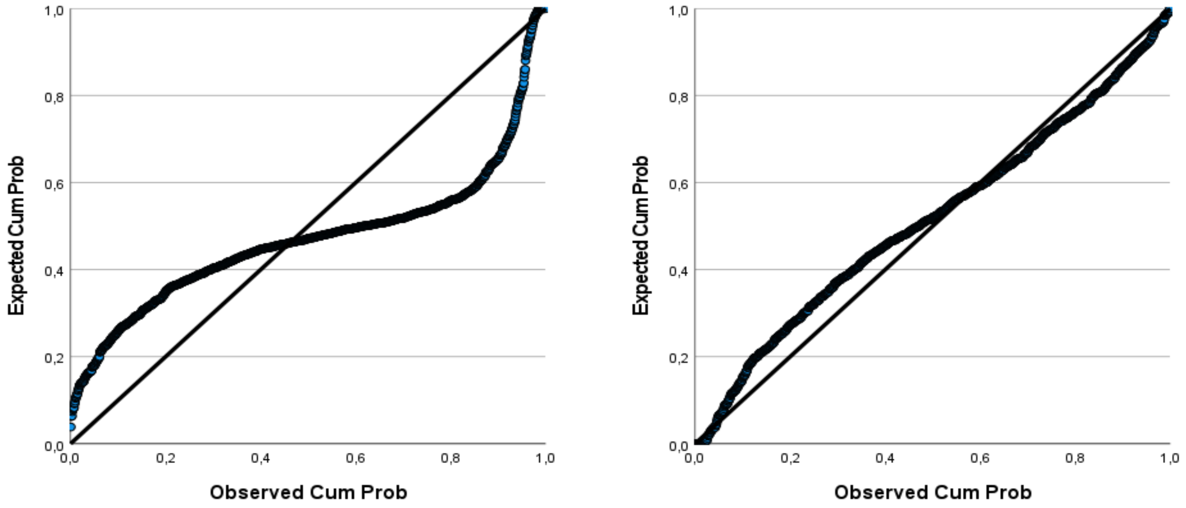

Figure 21. Normal P-P Plot of Regression Standardized Residual for the normal electricity consumption variable on the left and the natural logarithm of consumption on the right.

The test for heteroskedasticity also confirms that the model on the regular electricity consumption values leads to a high level of heteroskedasticity. More specifically, the error terms are unequally distributed above and below zero for different levels of consumption and also increase in variability with higher levels of consumption as can be seen in Figure 22. Applying the natural logarithm to the consumption improves homoscedasticity, even though there are still some outliers for medium and very high levels of consumption. Here, the model runs into the same problems of unclear impact of pure and impure heteroskedasticity and lack of means to resolve this issue further. Thus, it is again important to keep the potential limitations in mind. Again, the majority of errors are fairly equally distributed and, as long as the potential bias of the model in the mentioned areas is kept in mind, the model can still be legitimately applied to project consumption.

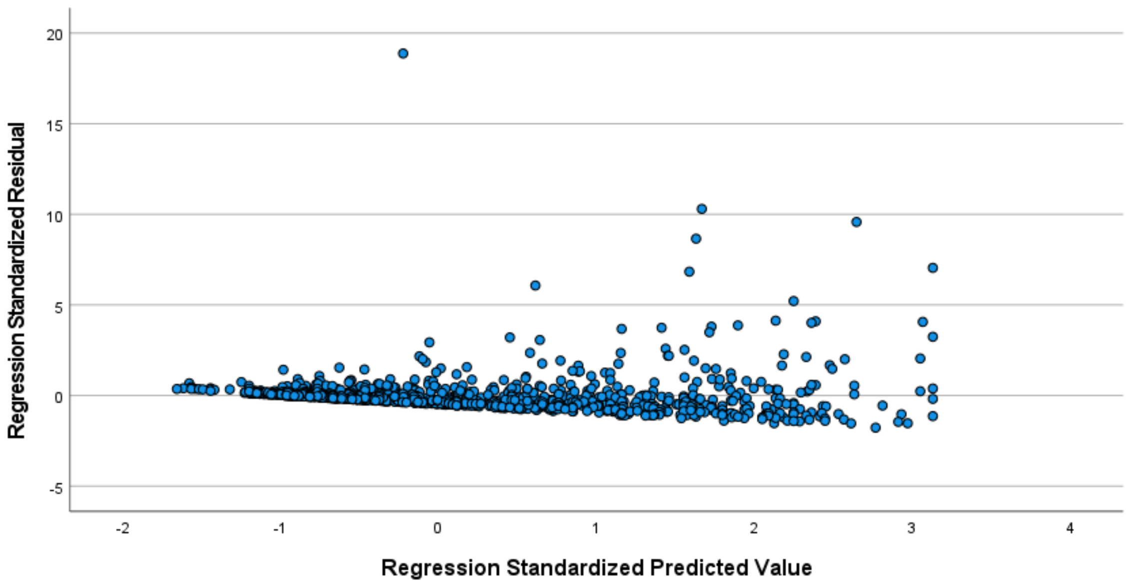

Figure 22. Scatterplot for the standardized residuals and the standardized predicted value calculated by the linear regression on electricity consumption.

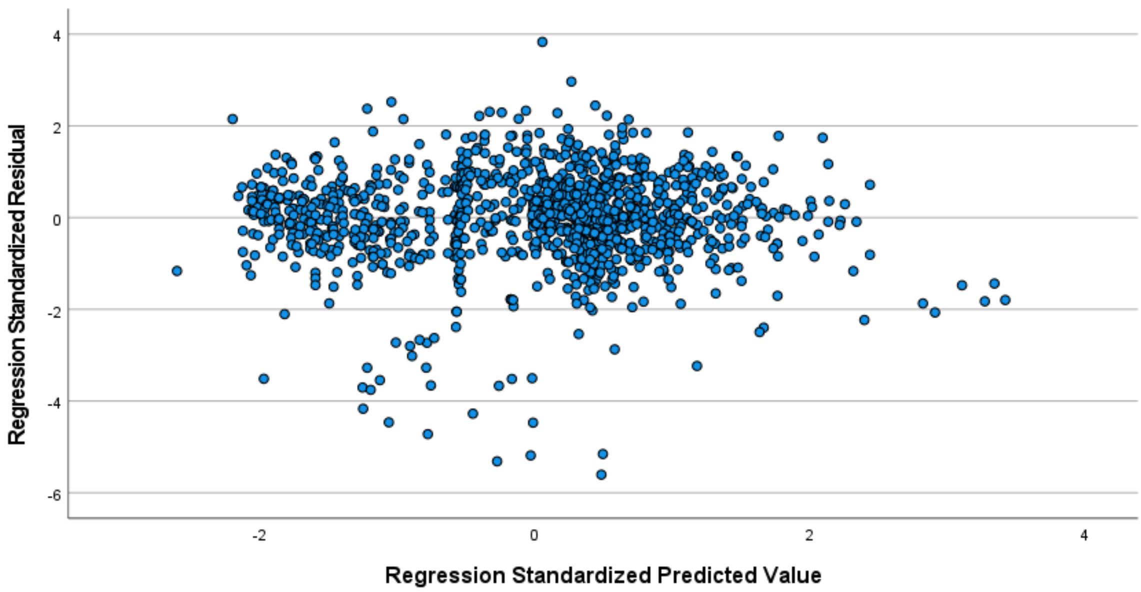

Figure 23. Scatterplot for the standardized residuals and the standardized predicted value calculated by the linear regression on the natural logarithm of electricity consumption.

9 Discussion of Findings, Limitations and Recommendations for Future Research

The research could help to identify predictors for EP uptake, performance and electricity consumption and confirm many qualitative findings or case-study level findings from literature on a quantitative, cross-country scale. To the knowledge of the author, a wider range of predictors and contexts have been analysed than in any previous study. Also, the attempt to derive an empirical model on the predictors of the level of electricity consumption for productive uses is new in literature. Accordingly, modelling the trajectory of enterprise uptake over performance up to estimating electricity consumption in one cohesive model is, to the knowledge of the other, unprecedented in current literature. Furthermore, the regressions on uptake, performance and electricity consumption in previous literature only aimed at understanding drivers and did not have the purpose to be used to derive projections for any other contexts. This required the consideration of previously unattended aspects, especially criteria of external validity in terms of representativeness and generalizability. The challenge of this study was to find the right balance of understanding bottom-up factors and translating them into variables which are available for all of SSA and generalizable for the wide, heterogenous group of microenterprises. Despite obvious challenges in finding these factors and the according data, it was possible to derive models which deliver reasonable projections for 2030 for the continent of SSA. Different modelling options were tested for their feasibility and quality to select the best modelling options. The report indicates the benefits and drawbacks of the discarded modelling options and explains which requirements need to be fulfilled in order to potentially make use of these alternative modelling approaches in future research. Obviously, also the final model is not free of flaws, which were diligently outlined in the report and will be summarized in the following sections to derive recommendations on how to tackle these issues for future research. Nevertheless, the models serve to set the frame for further studies to deepen the understanding of microenterprises in SSA and their interactions within different contexts. The study also gives a first idea about the order of magnitude of business activity, performance and electricity consumption of microenterprises in SSA.

Once implemented in IMAGE, it will also serve to give first insights into the prospective activity and electricity consumption of microenterprises and its implications for energy access planning. It will be investigated how the consideration of productive uses on top of residential demand might impact the cost and technologies in different policy and electrification scenarios.

In order to fully understand the extent to which the present models can be expected to contribute to a better understanding of the expected electricity demand for productive uses in 2030, the following sections indicate again the different limitations of the model. To give some structure to the limitations, they were sorted for conceptual limitations, data quality limitations, modelling limitations and statistical limitations.

9.1 Conceptual and normative limitations

As addressed in section 2, the desirability of electrification as well as increased microenterprise activity are discussed rather controversially in literature. If benefits from either are perceived by the concerned people depends on the given contexts. Seeing EP as mere potential consumers of electricity to make electrification worthwhile to energy providers can lead to paradoxical dynamic in which people might be incentivized to use and ultimately need electricity for productive uses so the energy providers can afford to entertain the connection, which might not have been needed or perhaps would have been sufficient with a lower load. This is a questionable concept of a win-win situation. This issue is not a limitation in the sense that it would render the study useless, of course, but should rather be understood as an appeal for caution so policy planning for the promotion of entrepreneurial activity and decisions about extent of supply are done carefully and critically. Both electricity supply and microenterprise uptake should not end up being seen as ends in themselves. Surrounding factors determine if and how electrification and business uptake should take place, so they contribute sustainably to the livelihoods of people (non-profitable electricity supply followed by lack of maintenance and deterioration of components, and rapid uptake and closure of businesses being the unsustainable scenario). Both, over- and underestimations of electricity demand are harmful for the concerned stakeholders.

Already when trying to define productive uses and microenterprises, one runs into severe problems as explained in section 2.1. Perhaps the universal definition is too broad, allowing too much heterogeneity to find a "narrow", linear explications. While concepts like the push/pull differentiation are useful, they are still clearly oversimplifying the real situation. However, the present definitions and concepts are commonly used across literature so the problem might not be in the definition perhaps simply in the heterogenous nature of microenterprises.

Concerning crop production, the projection was conducted based on the implicit assumption that there are no other barriers to processing other than access to electricity or that other issues will be solved until 2030. However, there are some noteworthy potential barriers and uncertainties, for example about the suitability of the selected crops for trade rather than subsistence use as well as suitability of the selected production processes for small scales, especially for sugar and oil production, which are quite energy intensive. In fact, interviewed farmers indicated openness to produce sugar beet but reported lack farming equipment and close-by processing facilities (Mandere et al., 2011).

9.2 Data and sample limitations

Many concerns have been addressed in the course of this report with respect to data availability and data quality. Many potentially important predictors could not be addressed due to measurability (how to quantify working morals, gender equality, corruption, etc.), availability in general (awareness and acceptance of productive uses, level of competition, public transportation, etc.) and availability for the projection (years of entrepreneurial experience, registration status, etc.). None of the sources used conducted the data collection for the precise purposes of assessing and projecting uptake, performance and electricity consumption of microenterprises, so the relevant information had to be pieced together from many different sources. Lack of data can lead to a generally lower share of explained variation, higher heteroskedasticity and confounding variable bias.

For the crop processing projections, the lack of data made it inevitable to make many generalized assumptions and compromises on the scale of facility, country scope and product had to be made. Taking into account different scales (or in general scales which are too large, as it had to be done for particularly for oil and sugar processing) is especially problematic, as different types of technologies can be associated, which might have different degrees of mechanization and electrification as well as different efficiencies. It is thus possible that a large-scale facility consumes more electricity per unit of output because more steps are electrified but it can also be that a large-scale facility consumes less because there will be economies of scale. Therefore, it cannot be said for sure to which extent the consideration of different scales impacts the electricity consumption estimates.

With the available data, some missing predictors could be approximated, however, this sometimes came to the cost of high levels of abstraction which obscure the insight into actual causality (e.g. state-level averages of literacy and household-head distributions, GNI PPP, etc.). Calculations based on approximations imposed further uncertainty on the variables (e.g. calculating electricity consumption based on cost in LCU). This undermines, to some extent, the intended bottom-up strategy of the PrElGen framework. In the case of the business type classification, the ISIC codes were found to be insufficiently accurate to do justice to the actual types of microenterprises in SSA.

Even though only data sources were used which were deemed trustworthy and which published their methodology, not all modalities of the data collection are known and often, It had to be accepted that the assumptions and exact calculations behind data were not fully clear (e.g. the calculation of the wealth quintiles). This means accepting some uncertainty about data quality, internal validity and representativeness. Furthermore, the EP Surveys included some questions to the interviewer about perceptions of trustworthiness of the respondents' answers and some imperfections had to be accepted (e.g. estimates instead of values form official records) in order to have a large enough sample size for building the model (see Table 12). Furthermore, the survey documentations often reported problems with recruitment as their sample frames were usually rather unreliable.

Table 12. Interviewers' perceptions about the quality of answers given my respondents in the WB EP Survey.

	Responses regarding opinions and perceptions
Not provided	0.2%
Truthful (reflect real opinions)	61.5%
Somewhat truthful	37.1%
Not truthful	1.2%
	Responses regarding figures
Are taken directly from establishments' records	15.2%
Are estimates computed with some precision	84.8%

There are some concerns about representativeness of the samples since only one country was investigated for the uptake and connection model and the selection of countries from the EP Surveys was dictated more by the availability of data rather than a diligent selection process. A further limitation was also imposed due to the calculation of weights in the HH Survey, which made samples on EP and community level non-representative. This might bias the results in regression analyses. Even though it can be said that different strata are all represented and only their distributions might differ from the population, some caution is still necessary. Furthermore, the application of weights was not possible in the EP Surveys because they were not given for all countries, rendering the sample not representative as well.

Perhaps the main representativeness-aspect to consider is the sampling strategy of the WB EP Surveys, which did not aim to represent all types of microenterprises but had a clear focus on urban areas, thereby also overrepresenting EP in largely wealthy areas and formal EP. Having the knowledge about this condition, it is possible to work with it as an assumption. It could imply that performance and electricity consumption could be overestimated as these businesses emerged under above-average favourable conditions. On the other hand, it could also be a legitimate assumption that in 2030, more favourable business operation conditions should be present on a broader scale so more businesses like the one in the sample can be expected. However, it is important to clearly communicate this assumption, as this direction of future development is uncertain and the projections for rural areas might be misrepresented.

Moreover, there is always the possibility of self-selection bias in surveys, which can be limited by careful sampling but not fully erased. Another problem is the accuracy of self-reported income and cost figures. There is a general tendency that income values are not reported or underreported, especially concerning high income levels and income in developing countries due to fear of tax investigation (Akpan et al., 2013). De Mel et al. (2009) for instance, show that people who are self-employed tend to under-report enterprise revenues by about 30%. Electricity consumption, on the other hand, especially future consumption, tends to be overestimated. Blodgett et al. (2017) found considerable discrepancies of an average of 15% between self-stated consumption volumes of households in Kenya and their actual consumption. On a study on short term estimation of past daily consumption levels, they even found a MAE of 426 Wh/day on an actual mean consumption of 113 Wh/day.

Next to the problem of generally missing data on certain variables, there was also the problem of missing values in the EP surveys for the two dependent variables, sales and electricity consumption. This is also a problem related to the bias in income reports. There is, however, no way of knowing the missing values for sure, so some lack of information had to be accepted.

Despite careful comparison, is also questionable how well the sample countries represent SSA, because of the ways in which countries differ amongst each other but also within. For example, cultural differences surely play a critical role but are often practically impossible to even measure.

9.3 Modelling limitations

The limitations classified as modelling limitations are closely linked to data issues but are more about the applicability and interpretability of data in the present models, rather than the quality or availability of specific variables.

One of those limitations is that for most variables (approx. 90%), only present or historical values are available because IMAGE only provides projections for some of their variables based on the SSPs. Since the main scenario, SSP2, assumes a moderate development with little changes compared to the present situation, the implicitly assumed stagnation of independent variables might not be too much of a break with the general scenario.

A certain limitation that results from lack of adequate data is that the exact causality between the dependent variable and the covariates might be statistically significant, but the underlying causality cannot be derived from mere statistics. Drawing on qualitative literature, it has been attempted to represent most causalities but non-linear relationships, confounding (missing) variables and interaction effects might bias the coefficients because additional qualitative information from the sample on these issues is missing. As the aim was to model stable supply in the connection and consumption models, the ambiguity of causality of some variables might mean that some of the predictors are actually modeling the supply rather than the demand for electricity, despite several precautions.

In general, the models, especially the one for electricity consumption, are quite simplified versions of reality, as they are trying to find a linear relationship between highly complex decisions and mechanisms. The insights into decision-making processes of households, time-dependence and specific electricity consumption patterns by different types of businesses (e.g. demand elasticity), had to be compressed to more superficial, simplified linear versions of reality. Attempts to deepening insights, e.g. by forming subsets, business type clusters, including interaction effects or factoring in time-dependency with panel data, were impaired mostly by the limited availability of required data. Accounting for these effects would also most likely improve the statistical quality of the overall model.

Finally, some variables which were on the entrepreneur level (phone ownership and education) had to be approximated in the projection with HH or individual averages while the distribution of these variables can be expected to be somewhat different amongst entrepreneurs as compared to the average population.

9.4 Statistical limitations

Statistical limitations are all issues faced concerning the quality criteria of the regression models. These are also largely symptoms of data quality and availability.

Overall, the models perform quite decently in explaining the deviation of observed values from the mean and the predictors show high levels of significance, considering the heterogeneity of EPs and HHs in terms of their wealth and locations. The R^2 of below 0.2 of the uptake model is of some concern but given that the distribution of projected probabilities seems largely reasonable, the final assessment on whether or not this model is fit for the projection will be seen in the actual projection results (see section 3.4 for strategies to evaluate the models based on the projections). Some further weaknesses are a tendency of the uptake model to overestimate EP presence and a tendency of underestimating values of the sales and consumption models. All of the regressions also have a natural tendency to regress to the middle and therefore the expected values underrepresent the extreme values observed in the sample. Furthermore, the linear regression models have some issues with heteroskedasticity and have larger error terms for medium and high values of the dependent variable, which might diminish the accuracy of the models.

9.5 Conclusion/Discussion of usefulness

In the context of all these limitations, the relevance of the third research question to ask, how much the model finally contribute as compared to the top down model., becomes very apparent. It will only be seen in the future, of course, how accurate the modelling results really are and even then, it cannot be said for sure in how far the developments took place due to the mechanisms suggested by the model. Given, however, the high relevance of microenterprises in the context of the SSA economies and their frequently stressed potential to make electrification financially sustainable make it clear that a mere random guess of consumption for productive uses, as is has been done before, does not do justice to the issue. The first step has been taken in the form of this research, to get a clearer picture of future developments of the electricity consumption and also the general activity of microenterprises in SSA, their drivers and methodologies to project these developments.

After having undertaken this challenge, the least that can be said is that the research provided an improved understanding of the difficulties in building an exact model for this purpose. However, given the diligent consideration of the choice of variables and the reasonable projection results, the given model should also provide a solid first start to build on. The many tested variables and modelling choices covered a variety of possible modelling pathways and the reporting was done in a way it is hoped to inform future research about the strengths and limitations of the different options. Despite the limited availability of data, quite some significant predictors could be identified, whose predictive power was stable even across different contexts. As hopefully, data on more predictors becomes available, and data quality improves the model can be extended and predictive power can be improved gradually.

Concerning its current applicability, it can be said that, even though the explained variances of the uptake and consumption model are relatively low, the model can still be expected to be useful for two reasons. First, the samples performed quite well in terms of projecting means/expected values for the dependent variables, with some tendency of overestimation in the uptake model and some underestimation in the sales and consumption models. Compared with literature, however, the projected values are still in line and overall plausible.

Secondly, for the purpose of the projection, projecting the accurate order of magnitude of the uptake of electricity demand for productive uses from microenterprises is already a large contribution to IMAGE, which previously estimated a rather arbitrary share of residential demand for productive uses. Especially since the main focus of the project is on projections of cost and technologies on the regionally aggregated scale (i.e. East, West, South and Central Africa), the limited accuracy in the variance is of secondary importance as long as the mean predicted values are realistic in scale. Having the additional transparency about the assumptions behind these projections helps to furthermore to interpret the results and put them into perspective. A sensitivity analysis intended to additionally consolidate the understanding of the robustness of these projections.

Even if the limitations turned out to render the model inapplicable under certain conditions, (e.g. in rural areas because the EP Surveys have such a strong bias for urban areas), this would already be additional, useful information for IMAGE and add some perspective, where previously no insight was available.

Moreover, the study could also contribute to the existing research on SSA microenterprises by confirming some findings from literature, add some context to ambiguous findings and discover some new predictors as well as explore some interaction effects of predictors quantitatively.

10 Future Research

Based on the findings and limitations, some suggestions can be derived for future research. In fact, one of the strengths of the research was the exploration of different modelling approaches which were previously not considered in the research on microenterprises in Africa.

10.1 Improving Data Quality

The more countries and the more contexts can be represented in the models, the better. But while the improvement of all modelling and projection input data is far from an easy fix, at least the key variables of the model could be improved by obtaining primary data with a sampling strategy and a questionnaire tailored to obtain the data relevant for the research. Especially knowing the actual electricity consumption rather than the expenditures for electricity would already be a great contribution. It would also be possible to get a deeper insight into the consumption elasticity of an enterprise, their vulnerability to outages an in how far their choice of enterprise was determined by the expected availability of electricity. It would also be important to know how much electricity they would like to consume at what time of the day to know the capacity of supply they need. It should also be ensured that the numbers are taken from actual records, rather than from memory or an estimate.

Another (additional) approach would be to improve the understanding of existing variables. Present literature already gives a few qualitative insights into the causal mechanisms and connections of predictors but oftentimes, these insights are guesses or not consistent over different studies. A well-sampled qualitative (pre-)study tailored specifically to the purpose of understanding motives behind enterprise uptake and electricity consumption would be of immense value to understand the weights as well as possible interactions of predictors. It could also allow to conclude how available data can be used to approximate omitted variables.

Furthermore, it could be attempted to improve the existing input data. At the moment, most variable values are projected to stay stable at the current levels, which is rather unlikely given the dynamic developments in SSA. While data availability is one reason that no development of these variables was projected, it was also beyond the scope of this research to investigate on, design and test possible developments. Future research could test in how far information for projections on different parameters (for example the share of female HHH) could develop or simply look at past values and model extrapolations of past trends based on the CAGR (compound annual growth rates) or similar assumptions. Of course, since SSA is one of the most dynamic regions in the world at the moment, any developments are highly uncertain. However, this makes the analysis of possible future pathways and the sensitivity to the projected uptake, performance and electricity consumption to these dynamics all the more interesting.

Data for crop processing could also be collected specifically from typical existing micro-scale crop-processing facilities. The possibility of having shared processing facilities could also be investigated, especially for the processes such as sugar and oil manufacturing, which might take too much energy for single farmers to pay off. A more elaborate selection of crops and products based on a tailored collection of primary data could be conducted in which issues such as a distinction between subsistence crops and cash crops and market potentials of different crops are assessed.

10.2 Improving the Empirical Models

Several ideas for further modelling approaches such as cluster and panel analyses, forming subsets or exploring interaction effects, etc. were already tested within the scope of this research. Most of the times, the attempts were promising at first but could not be executed fruitfully due to data limitations. If data availability could be improved, testing these models again could contribute a lot to the understanding of business activity and productive electricity uses, especially understanding the interaction of different predictors.

Especially research aimed at finding characteristics by which to distinguish business types effectively and bringing some structure into the existing heterogeneity can be expected to improve the accuracy of prediction models immensely.

Perhaps more transformations of variables can also improve the relationships. In the present study, it was attempted to refrain from data transformations as much as possible in this study to maintain transparency. It is possible, however, that some transformations could improve the accuracy of the model because relationships between the dependent and independent variables are more accurately represented by non-linear models.

10.3 Further Applications of the Model

Finally, the research findings can, to some extent, also be applied to project-scale electrification planning when more specific projections are needed but also more local, specific data is available. The models could be extended by implementing the variables which were only available for the samples but not for all of SSA. Some of these potentially useful variables are the status of registration, hours of operation per week, number of full-time and part-time employees or the age of the company. This might even allow to test some of the discarded modelling ideas.

11 Conclusion

In conclusion, the study managed to achieve its target to develop a model which can estimate the uptake and electricity consumption of microenterprises in SSA despite the limitations of data availability, quality and the lack in current literature of a generalizable conceptual understanding of the nature of micro-scale enterprises and their electricity consumption patterns.

In achieving its research objective, the study added to the existing body of research on microenterprise activity by confirming some predictors for their uptake, performance and electricity consumption found in present literature on an unprecedented empirical scale and range of contexts, exploring explanatory effects and uncovering some new predictors. Furthermore, several other approaches which were tested on the path to distilling the final model, are reported on with respect to their strengths and weaknesses, hoping to inspire and encourage future research. However, in acknowledging the vast range of limitations faced in the modelling process, the study also gave a structured overview of its weaknesses and how to take them into account for the interpretation of results. The report also indicated pathways to tackle these limitations and improve the model in the course of future research.

The study further met its objective of adding the component of productive uses to the IMAGE model, which can now give a better understanding of the order of magnitude of electricity demand for productive uses. It also provides an understanding for the underlying assumptions and predictors which are factored in (as well as those which could not be factored in) to determine the consumption for productive uses, which was previously only determined in a top-down manner with no insight into the underlying drivers or the expectable scale. As research proceeds and data availability and quality increase, it is hoped that the accuracy of the model will also increase. This bottom-up approach initiated by the PrElGen framework is hoped to improve the precision and transparency of electricity projections, which sets IMAGE apart from other electrification models known to the author. Ultimately, the improved methodology is hoped to better inform policy makers and plan the further electrification in SSA in a socially, economically and ecologically sustainable way.

12 References

Abidjan.net. (2020). *Électricité*. https://news.abidjan.net/h/595026.html

Abubakar, M. S., Umar, B., & Ahmad, D. (2010). Energy use patterns in sugar production: A case study of savannah sugar company, Numan, Adamawa State, Nigeria. *Journal of Applied Sciences Research*, *6*(4), 377–382.

Ackah, C. (2011). *NONFARM EMPLOYMENT AND INCOMES IN RURAL GHANA 1*. https://doi.org/10.1002/jid.1846

ADA. (2016). *SMALL AND GROWING BUSINESSES IN AFRICA: PROFILES, SUCCESSES AND CHALLENGES*.

Ahiduzzaman, M., & Islam, A. K. M. S. (2009). Energy Utilization and Environmental Aspects of Rice Processing Industries in Bangladesh. *Energies*, *2*, 134–149. https://doi.org/10.3390/en20100134

Aithnard, P.-H. (2014). *Middle Africa Insight Series | Power Fully charged : Key dynamics in Middle Africa ' s Power Sector in 2014* (Issue January). https://www.ecobank.com/upload/20140130120637783422UtM8wTMzbm.pdf

Ajala, A. S., & Gana, A. (2015). Analysis of Challenges Facing Rice Processing in Nigeria. *Journal of Food Processing*, *2015*, 1–6. https://doi.org/10.1155/2015/893673

Akpan, U., Essien, M., & Isihak, S. (2013). *Energy for Sustainable Development The impact of rural electri fi cation on rural micro-enterprises in*. *17*, 504–509.

Aliu, S. A., Onochie, U. P., Itabor, N. A., & Adingwupu, A. C. (2018). Energy audit of a flour mill plant: A case study of Crown Flour Mill PLC. *Nigerian Research Journal of Engineering and Environmental Sciences*, *3*(1), 345–358.

Alsos, G. A., Carter, S., & Ljunggren, E. (2013). *ENTREPRENEURIAL FAMILIES AND HOUSEHOLDS*.

Ambrose, D. C. P., Annamalai, S. J. K., Naik, R., Dubey, A. K., & Chakraborthy, S. (2017). Performance studies on millet processing machinery for tribal livelihood promotion. *Journal of Applied and Natural Science*, *9*(3), 1796–1800. https://doi.org/10.31018/jans.v9i3.1441

Aterido, R., & Hallward-Driemeier, M. (2010). The Impact of the Investment Climate on Employment Growth Does Sub-Saharan Africa Mirror Other Low-Income Regions? *The World Bank Policy Research Working Paper*, *February*, 1–42. http://documents.worldbank.org/curated/en/965531468192859171/pdf/WPS5218.pdf

Atiku, A. A., Aviara, N. A., & Haque, M. A. (2004). *PERFORMANCE EVALUATION OF A BAMBARA GROUND NUT SHELLER*. *VI*, 1–18.

Balasubramanian, S. (2015). *Abstracts Under the aegis of the*. *September 2013*. https://doi.org/10.13140/RG.2.1.4318.7365

Banerjee, S. G., Nash, J., Malik, K., Tipping, A., & Besnard, J. (2017). *Double Dividend : Power and Agriculture Nexus in Sub-Saharan Africa*.

Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2004). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyh299

Bassey, M. W., & Schmidt, O. G. (1989). *Abrasive-disk dehullers in Africa from research to dissemination* (G. C. R. Croome (ed.)). International Development Research Centre.

Beguería, S., & Vicente Serrano, S. M. (2016). *SPEIbase v. 2.4.*

Benedikter, S., Waibel, G., Birtel, S., Bui, C. T., & Tran, B. T. (2013). *Local Entrepreneurship in Vietnam's Rural Transformation. A Case Study from the Mekong Delta*. *September 2016*, 1–92. https://mpra.ub.uni-muenchen.de/49866/

Bensch, G., Kreibaum, M., Mbegalo, T., Peters, J., & Wagner, N. (2016). Materialien The Status of Energy Access in Three Regions of Tanzania Baseline report for an urban grid upgrading. *RWI Materialien*, *111*.

Berisha, G., & Pula, J. S. (2015). *Defining Small and Medium Enterprises : a critical review Defining Small and Medium Enterprises : a critical review*. *March*.

Bernard, T. (2012). Impact analysis of rural electrification projects in Sub-Saharan Africa. *World Bank Research Observer*, *27*(1), 33–51. https://doi.org/10.1093/wbro/lkq008

Bertheau, P., Oyewo, A. S., Cader, C., Breyer, C., & Blechinger, P. (2017). Visualizing national electrification scenarios for sub-saharan African countries. *Energies*, *10*(11), 1–20. https://doi.org/10.3390/en10111899

Blodgett, C., Dauenhauer, P., Louie, H., & Kickham, L. (2017). Accuracy of energy-use surveys in predicting rural mini-grid user consumption. *Energy for Sustainable Development*, *41*, 88–105. https://doi.org/10.1016/j.esd.2017.08.002

Blodgett, C., Moder, E., Kickham, L., & Leaf, H. (2016). *Powering Productivity. early Insights into Mini Grid Operations in Rural Kenya.*

Booth, S., Li, X., Baring-Gould, I., Kollanyi, D., Bharadwaj, A., & Weston, P. (2018). Productive Use of Energy in African Micro-Grids : Technical and Business Considerations. In *Usaid, Nrel* (Issue August). www.nrel.gov/publications.%0Ahttps://www.nrel.gov/docs/fy18osti/71663.pdf

Brew-Hammond, A. (2010). Energy access in Africa: Challenges ahead. *Energy Policy*, *38*(5), 2291–2301. https://doi.org/10.1016/j.enpol.2009.12.016

Briceño-Garmendia, C., & Domínguez-Torres, C. (2010). Kenya ' s Infrastructure : A Continental Perspective. *Development*, *March*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1792254

Briceño-Garmendia, C. M., & Benitez, D. A. (2011). *Cape Verde's Infrastructure. A Continental Perspective*. World Bank.

Bruhn, M., Hommes, M., Khanna, M., Singh, S., Sorokina, A., & Wimpey, J. S. (2017). MSME Finance Gap: Assessment of the Shortfalls and Opportunities in Financing Micro, Small and Medium Enterprises in Emerging Markets,. In *MSME Finance Gap*. https://doi.org/10.1596/28881

Bungane, B. (2016). *Zimbabwe: ZERA approves electricity tariff increase*. https://www.esi-africa.com/regional-news/southern-africa/zimbabwe-zera-approves-electricity-tariff-increase/

Cabraal, A., Barnes, D. F., & Agarwal, S. G. (2005). Productive uses of energy for rural development. In *Annual Review of Environment and Resources* (Vol. 30, Issue March, pp. 117–144). https://doi.org/10.1146/annurev.energy.30.050504.144228

Canagarajah, S., Newman, C., & Bhattamishra, R. (2001). Non-farm income, gender, and inequality: evidence from rural Ghana and Uganda. *Food Policy*, *26*, 405–420. www.elsevier.com/locate/foodpol

Ceesay, M. L. S. (2012). *General overview of The Gambia's Electricity Market and Regulatory Framework for Renewable Energy*. PURA. https://pubs.naruc.org/pub.cfm?id=5385E484-2354-D714-51D3-F44DCFDFFFF9

CfA. (2005). *Our Common Interest report ofthe Commision for Africa.* https://doi.org/10.1017/CBO9781107415324.004

Chen, H., & Chen, G. Q. (2011). *Energy cost of rapeseed-based biodiesel as alternative energy in China*. *36*, 1374–1378. https://doi.org/10.1016/j.renene.2010.11.026

Chin, R., Matlashewski, N., & Swan, K. (2012). *Development of a Little Millet Mill*.

Chouhan, P., & Chandrakar, A. (2014). Performance Enhancement of Sugar Mill by Alternate Cooling System for Condenser. *IOSR Journal of Mechanical and Civil Engineering*, *11*(5), 18–25.

Christiaensen, L., & Demery, L. (2017). Agriculture in Africa: Telling Myths from Facts. In L. Christiaensen & L. Demery (Eds.), *Agriculture in Africa: Telling Myths from Facts*. The World Bank. https://doi.org/10.1596/978-1-4648-1134-0

CHU, H. M., BENZING, C., & MCGEE, C. (2007). GHANAIAN AND KENYAN ENTREPRENEURS: A COMPARATIVE

ANALYSIS OF THEIR MOTIVATIONS, SUCCESS CHARACTERISTICS AND PROBLEMS. *Journal of Developmental Entrepreneurship*, *12*(03), 295–322. https://doi.org/10.1142/S1084946707000691

CIESIN. (2013). *Global Roads Open Access Data Set (gROADS), v1: Global Roads | SEDAC*. https://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1

Costa, R., & Rijkers, B. (2011). Gender and Rural Non-farm Entrepreneurship. *World Development.*, *40*(12), 2411–2426.

Cronk, R., & Bartram, J. (2018). Environmental conditions in health care facilities in low- and middle-income countries: Coverage and inequalities. *International Journal of Hygiene and Environmental Health*, *221*(3), 409–422. https://doi.org/10.1016/j.ijheh.2018.01.004

Da, G., Dufour, D., Giraldo, A., Moreno, M., Tran, T., & Velez, G. (2013). *Cottage Level Cassava Starch Processing Systems in Colombia and Vietnam*. 2213–2222. https://doi.org/10.1007/s11947-012-0810-0

Dagnachew, A. G., Hof, A. F., Lucas, P. L., & van Vuuren, D. P. (2020). Scenario analysis for promoting clean cooking in Sub-Saharan Africa: Costs and benefits. *Energy*, *192*. https://doi.org/10.1016/j.energy.2019.116641

Dagnachew, A. G., Lucas, P. L., Hof, A. F., Gernaat, D. E. H. J., de Boer, H. S., & van Vuuren, D. P. (2017). The role of decentralized systems in providing universal electricity access in Sub-Saharan Africa – A model-based approach. *Energy*, *139*, 184–195. https://doi.org/10.1016/j.energy.2017.07.144

Dagnachew, A. G., Lucas, P. L., Hof, A. F., & van Vuuren, D. P. (2018). Trade-offs and synergies between universal electricity access and climate change mitigation in Sub-Saharan Africa. *Energy Policy*, *114*(November 2017), 355–366. https://doi.org/10.1016/j.enpol.2017.12.023

Daioglou, V., van Ruijven, B. J., & van Vuuren, D. P. (2012). Model projections for household energy use in developing countries. *Energy*, *37*(1), 601–615. https://doi.org/10.1016/j.energy.2011.10.044

de Mel, S., McKenzie, D. J., & Woodruff, C. (2009). Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics*, *88*(1), 19–31. https://doi.org/10.1016/j.jdeveco.2008.01.007

Dozier, W. A., Hanna, W., & Behnke, K. (2005). *Grinding and Pelleting Responses of Pearl Millet-Based Diets 1*. 2–7.

ECA. (2009). *The Potential of Regional Power Sector Integration. South African Power Pool (SAPP) | Transmission % Trading Case Study. Regional Power Sector Integration:Lessons From Global Case Studies And A Literature Review. ESMAP Briefing Note 004/10.* http://documents1.worldbank.org/curated/en/126211468323330432/pdf/773070v110ESMA0h0African0Power0Pool.pdf

EEC. (2018). *Electricity Tariffs from Swaziland Electricity Company*. http://www.eec.co.sz/myaccount/tariffs/index.php

Efobi, U. R., Beecroft, I., & Atata, S. N. (2019). *Female Access and Rights to Land , and Rural Non - farm Entrepreneurship in Four African Countries*. *31*(2), 179–189.

EIA. (2020). *The Annual Energy Outlook explores long-term energy trends in the United States*. www.eia.gov/aeo

Elvidge, C. D., Baugh, K., Zhizhin, M., Chi Hsu, F., & Ghosh, T. (2017). Feng Chi Hsu & Tilottama Ghosh (2017) VIIRS night-time lights. *International Journal of Remote Sensing*, *38*(21), 5860–5879. https://doi.org/10.1080/01431161.2017.1342050

ENEO. (2015). *Eneo Cameroon S.A.: The energy of Cameroon - MyEasyLight, your e-agency*. https://www.eneocameroon.cm/index.php/en/eneo-client-guide/eneo-annual-report-2014-managing-my-account

Eshetu, A., Kamil, A., Ashenir, T., Jemal, N., & Gelgelo, K. (2017). Regional Review Workshop on Completed Research Activities. *Proceedings of Review Workshop on Completed Research Activities of Agricultural Engineering Research Directorate*.

ESI Africa. (2019). *Tariffs renegotiation with IPPs will unsettle green energy investments.* https://www.esi-africa.com/industry-sectors/renewable-energy/tariffs-renegotiation-with-ipps-will-unsettle-green-energy-investments/

Etim, E., & Iwu, C. G. (2019). A descriptive literature review of the continued marginalisation of female entrepreneurs in sub-Saharan Africa. *International Journal of Gender Studies in Developing Societies*, *3*(1), 1. https://doi.org/10.1504/ijgsds.2019.096755

Falchetta, G. (n.d.). *Multi-sectoral bottom-up assessment of community electrification to improve energy access decisions*.

Falchetta, G., & Noussan, M. (2019). *Interannual Variation in Night-Time Light Radiance Conditional on Income-Level and Region*. https://doi.org/10.3390/en12030456

Fishbein, R. E., Sanghvi, A. P., Energy, A., & World, U. (2003). Survey of Productive Uses of Electricity in Rural Areas. *World*, *April*, 1–51.

Florczyk, A. J., Corbane, J., Ehrlich, C., Freire, D., Kemper, S., Maffenini, T., Melchiorri, L., Pesaresi, M., Politis, M., Schiavina, P., Sabo, M., & Zanchetta, F. (2019). *GHSL Data Package 2019* (Issue July). https://doi.org/10.2760/290498

Fore, S. R., Porter, P., & Lazarus, W. (2011). *Net energy balance of small-scale on-farm biodiesel production from canola and soybean*. *5*. https://doi.org/10.1016/j.biombioe.2011.02.037

Foster, V., & Briceño-Garmendia, C. (2010). *Africa ' s Infrastructure : A Time for Transformation*.

Foster, V., & Steinbuks, J. (2009). Paying the Price for Unreliable Power Supplies: In-House Generation of Electricity by Firms in Africa. *World Bank Policy Research Working Paper*, *4913*.

Fox, L., & Sekkel Gaal, M. (2008). Working Out of Poverty. Job Creation and the Quality of Growth in Africa. In *Directions in Development. Poverty.* https://doi.org/10.1596/978-0-8213-7442-9

FRES. (2013). *Foundation Rural Energy Services Socio-economic Impact Assessment of* (Issue July). FRES (Foundation RuralEnergy Services. PricewaterhouseCoopers PwC.

Garcia-Valentin, A., Mestres, C. A., Bernabeu, E., Bahamonde, J. A., Martín, I., Rueda, C., Domenech, A., Valencia, J., Fletcher, D., Machado, F., & Amores, J. (2015). *Validation and quality measurements for EuroSCORE and EuroSCORE II in the Spanish cardiac surgical population: a prospective, multicentre study †*. https://doi.org/10.1093/ejcts/ezv090

Gebre-Egziabher, T. (2009). The developmental impact of Asian drivers on Ethiopia with emphasis on small-scale footwear producers. *World Economy*, *32*(11), 1613–1637. https://doi.org/10.1111/j.1467-9701.2009.01252.x

Gibson, J., & Olivia, S. (2010). The effect of infrastructure access and quality on non-farm enterprises in rural Indonesia. *World Development*, *38*(5), 717–726. https://doi.org/10.1016/j.worlddev.2009.11.010

Glejser, H. (1969). A New Test for Heteroskedasticity. In *Source: Journal of the American Statistical Association* (Vol. 64, Issue 325).

GlobalPetrolPrices.com. (2020). *Mozambique electricity prices, December 2019 | GlobalPetrolPrices.com*. https://www.globalpetrolprices.com/Mozambique/electricity_prices/

Golumbeanu, R., & Barnes, D. (2013). *Connection Charges and Electricity Access in Sub-Saharan Africa The World Bank Africa Region Sustainable Development Network*. *June*. http://econ.worldbank.org.

Goyal, S. K., & Agrawal, A. K. (2008). A study of energy audit in rice processing machines. *Prog. Agile.*, *8*(January 2008), 34–38.

Goyal, S. K., Chandra, S., & Chandra, S. (2010). *Energy Use Pattern in Rice Processing Plants* (pp. 309–311). Society for Recent Development in Agriculture.

Green, C. R., Sodiki, J. I., & Nkoi, B. (2019). *Energy Audit of a Wheat Processing Plant in Port*. *4*(1), 69–74.

Grimm, M., & Hartwig, R. (2012). *How much does Utility Access matter for the Performance of Micro and Small*.

1–31.

Haggblade, S., Hazell, P. B. R., & Reardon, T. (2007). *Transforming the Rural Nonfarm Economy. Opportunities and Threats in the Developing World.* International Food Policy Research Institute; The Johns Hopkins University Press.

Hallberg, K. (2000). A Market-Oriented Strategy for Small and Medium Scale Enterprises. In *A Market-Oriented Strategy for Small and Medium Scale Enterprises*. The World Bank. https://doi.org/10.1596/0-8213-4727-6

Harsdorff, M., & Bamanyaki, P. (2009). *IMPACT ASSESSMENT OF THE SOLAR ELECTRIFICATION OF MICRO ENTERPRISES , HOUSEHOLDS AND THE DEVELOPMENT* (Issue September).

Herrington, M., & Kelley, D. (2013). African Entrepreneurship. Sub-Saharan African Regional Report. In *African Affairs*.

Huenteler, J., Dobozi, I., Balabanyan, A., & Banerjee, S. (2017). *Cost Recovery and Financial Viability of the Power Sector in Developing Countries: A Literature Review*. *December*, 1–54. https://doi.org/10.1596/1813-9450-8287

IEA. (2019a). Africa Energy Outlook 2019 – Analysis Scenarios. *World Energy Outlook Special Report*, 288. https://www.iea.org/reports/africa-energy-outlook-2019#energy-access%0Ahttps://www.iea.org/reports/africa-energy-outlook-2019#africa-case

IEA. (2019b). *Defining energy access: 2019 methodology*. https://www.iea.org/articles/defining-energy-access-2019-methodology

IFC. (2020). *IFC's Definitions of Targeted Sectors*. https://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/financial+institutions/priorities/ifcs+definitions+of+targeted+sectors

ILO, UNIDO, & FAO. (1984). *Small-Scale Maize Milling*.

Ion, I. V, & Popescu, F. (2017). *Improving the Energy Balance in a Sunflower Oil Mill*. *Eurostat 2015*, 164–171.

Jekayinfa, S. O., & Olajide, J. O. (2007). *Analysis of energy usage in the production of three selected cassava-based foods in Nigeria*. *82*, 217–226. https://doi.org/10.1016/j.jfoodeng.2007.02.003

Jensen, S., Morin, B., Jensen, A. S., & Morin, B. (2015). *Energy and the environment in beet sugar production Energy and the environment in beet sugar production*.

Jones, S., Gibbon, P., Jones, S. A. M., & Gibbon, P. (2011). *Developing Agricultural Markets in Sub-Saharan Africa : Organic Cocoa in Rural Uganda Developing Agricultural Markets in Sub-Saharan Africa : Organic Cocoa in Rural Uganda*. *0388*. https://doi.org/10.1080/00220388.2011.579107

Kariuki, D., Kimuyu, P., & Nyangena, D. (2016). Rural Electrification and Microenterprises Performance: Some lessons from Muranga County, Kenya. *International Journal of Economics*, *1*(1), 31–45.

Kartika, I. A., Pontalier, P. Y., & Rigal, L. (2010). *Twin-screw extruder for oil processing of sunflower seeds : Thermo-mechanical pressing and solvent extraction in a single step*. *32*, 297–304. https://doi.org/10.1016/j.indcrop.2010.05.005

Kirubi, C., Jacobson, A., Kammen, D. M., & Mills, A. (2009). Community-Based Electric Micro-Grids Can Contribute to Rural Development: Evidence from Kenya. *World Development*, *37*(7), 1208–1221. https://doi.org/10.1016/j.worlddev.2008.11.005

Kooijnman-van Dijk, A. L. (2008). *The power to produce. The Role of energy in poverty reduction through small scale enterprises in the Indian Himalayas.* Universiteit Twente.

Korkovelos, A., Khavari, B., Sahlberg, A., Howells, M., & Arderne, C. (2019). The role of open access data in geospatial electrification planning and the achievement of SDG7. An onsset-based case study for Malawi. *Energies*, *12*(7). https://doi.org/10.3390/en12071395

Kuwonu, F. (2019). *Black-eyed peas: A taste of Africa in the Americas | Africa Renewal*.

https://www.un.org/africarenewal/magazine/december-2019-march-2020/black-eyed-peas-taste-africa-americas

LAC. (2008). *Government Gazette of The Republic of Namibia. No. 4102*.

Lal, R. R., & Verma, P. (2007). *Post-Harvest Management of Pulses*. Indian Institute of Pulses Research.

Lawrence, K. A. (2011). Soybean: Africa's Potential Cinderella Food Crop. *Soybean - Biochemistry, Chemistry and Physiology*, *June*. https://doi.org/10.5772/15527

Lecoque, D., & Wiemann, M. (2015). The Productive Use of Renewable Energy in Africa. In *European Union Energy Initiative Partnership Dialogue Facility (EUEI PDF)*. http://ruralelec.org/38.0.html

Lenz, L., Munyehirwe, A., Peters, J., & Sievert, M. (2017). Does Large-Scale Infrastructure Investment Alleviate Poverty? Impacts of Rwanda's Electricity Access Roll-Out Program. *World Development*, *89*, 88–110. https://doi.org/10.1016/j.worlddev.2016.08.003

Lipton, M. (1977). *Urban Bias in World Development* (Vol. 73, Issue 1).

Little, I. M. D., Mazumdar, D., & Page, J. M. (1987). Small Manufacturing Enterprises. In *A World Bank Research Publication*. http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=6774943&site=ehost-live

Loening, J., & Lane, W. L. (2007). *Tanzania: Pilot Rural Investment Climate Assessment. Stimulating Nonfarm Microenterprise Growth*.

Lombardi, F., Balderrama, S., Quoilin, S., & Colombo, E. (2019). *Generating high-resolution multi-energy load pro fi les for remote areas with an open-source stochastic model*. *177*.

Lynn, H. S. (2003). Suppression and Confounding in Action. *American Statistician*, *57*(1), 58–61. https://doi.org/10.1198/0003130031090

M. Soppimath, V., & G. Hudedmani, M. (2017). Energy Audit and Energy Management in the Sugar Industry. *MATTER: International Journal of Science and Technology*, *3*(1), 10–26. https://doi.org/10.20319/mijst.2017.31.1026

Maleko, G. C. (2005). *" Impact of Electricity Services on Microenterprise in Rural Areas in Tanzania " By a Thesis Submitted for the Award of Master of Environmental Business Administration (Environmental and Energy Management- Mba) Department of Energy and Sustainable Devel*. *December*.

Mandere, N., Anderberg, S., Armah, F. A., & Abaya, S. W. (2011). Assessing the contribution of alternative agriculture to poverty reduction and employment creation: A case study of sugar beet cultivation in Kenya. *African Journal of Agricultural Research*, *6*(2), 440–450. https://doi.org/10.5897/AJAR09.050

Mapako, M., & Mbewe, A. (2013). *Renewables and Energy for Rural Development in Sub-Saharan Africa*. Zed Books Ltd. https://books.google.nl/books?id=ZBNjDgAAQBAJ

Mapako, Maxwell, & Prasad, G. (2007). Rural electrification in Zimbabwe reduces poverty by targeting income-generating activities. *CSIR Natural Resources and the Environment, and Energy Research Centre, University of Cape Town*, 1–6.

Mawejje, J., Munyambonera, E., & Bategeka, L. (2012). *Uganda's Electricity Sector Reforms and Institutional Restructuring.*

Mayer-Tasch, L., Mukherjee, M., & Reiche, K. (2013). Productive Use of Energy – PRODUSE Measuring Impacts of Electrification on Small and Micro-Enterprises. *Report*, 156. http://www.produse.org/imglib/downloads/PRODUSE_study/PRODUSE Study_Full Text.pdf

Mead, D. C., & Liedholm, C. (1998). The dynamics of micro and small enterprises in developing countries. *World Development*, *26*(1), 61–74. https://doi.org/10.1016/S0305-750X(97)10010-9

Mentis, D., Howells, M., Rogner, H., Korkovelos, A., Arderne, C., Zepeda, E., Siyal, S., Taliotis, C., Bazilian, M., De Roo, A., Tanvez, Y., Oudalov, A., & Scholtz, E. (2017). Lighting the World: the first application of an open source, spatial electrification tool (OnSSET) on Sub-Saharan Africa. *Environmental Research Letters*, *12*(8). https://doi.org/10.1088/1748-9326/aa7b29

Moyo, B. (2013). *Power infrastructure quality and manufacturing productivity in Africa : A fi rm level analysis*. *61*, 1063–1070.

Muthoni, M. P. (2019). *Socio- Economic Impact Of Electrification On Micro And Medium-Sized Enterprises In Kibera Slum, Nairobi*. *October*.

Nagler, P., & Naudé, W. (2017). *Non-farm entrepreneurship in rural sub-Saharan Africa : New empirical evidence*. *67*, 175–191.

Nautiyal, P. C. (2002). *GROUNDNUT Post-harvest Operations*.

NAWEC. (2020). *Service Tariffs – NAWEC*. http://nawec.gm/?page_id=203

Neelsen, S., & Peters, J. (2011). *Energy for Sustainable Development Electricity usage in micro-enterprises — Evidence from Lake Victoria , Uganda ☆*. *15*, 21–31. https://doi.org/10.1016/j.esd.2010.11.003

Neter, J., Wasserman, W., & Kutner, M. H. (1990). Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design. *Applied Linear Statistical Models*.

Nkegbe, P. K. (2018). *Agriculture Commercialization in Ghana : November*.

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., & van Vuuren, D. P. (2014). A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400. https://doi.org/10.1007/s10584-013-0905-2

Okafor. (2016). *Electricity Tariff for 2012 - Nigeria Technology Guide*. NaijaTechGuide. https://www.naijatechguide.com/2012/06/new-electricity-tariff-for-2012.html

Okpara, J. O., & Wynn, P. (2007). Determinants of small business growth constraints in a sub-Saharan African economy. *SAM Advanced Management Journal*, *72*(2), 24–36. https://go-gale-com.proxy.library.uu.nl/ps/i.do?p=AONE&sw=w&issn=07497075&v=2.1&it=r&id=GALE%7CA166537560&sid=googleScholar&linkaccess=fulltext

Olaoye, O. S., Adefajo, A., & Ekundayo, S. O. (2020). Energy analysis of a wheat processing plant in Nigeria. *ARPN Journal of Engineering and Applied Sciences*, *9*(January 2014), 1586–1591.

Olawale, F., & Garwe, D. (2010). *Obstacles to the growth of new SMEs in South Africa : A principal component analysis approach*. *4*(May), 729–738.

Opoku, A., & Tabil, L. G. (2006). *Effect of microwave drying on grinding and particle size distribution of green field pea*. *December*. https://doi.org/10.13031/2013.21206

Owoo, N., & Naudé, W. (2014). Non-Farm Enterprise Productivity and Spatial Autocorrelation in Rural Africa: Evidence from Ethiopia and Nigeria. *IZA Discussion Papers*, *March 2014*.

Oxfam. (2017). *Inequality in Nigeria. Exploring the Drivers.*

PBL. (2014). *Integrated Assessment of Global Environmental Change with IMAGE 3.0. Model description and policy applications.* PBL Netherlands Environmental Assessment Agency.

Peters, J., Sievert, M., Toman, M. A., & Leibniz, R. W. I. (2019). *Rural electrification through mini-grids : Challenges ahead*. *132*(December 2018), 27–31.

Pippo, W. A., & Luengo, C. A. (2016). *Sugarcane energy use : accounting of feedstock energy considering current agro-industrial trends and their feasibility*. *April 2014*. https://doi.org/10.1186/2251-6832-4-10

PURA. (2008). *PURA ANNUAL REPORT*. http://www.pura.gm/wp-content/uploads/2018/01/Annual-Report-2008.pdf

Rao, N. D. (2013). *Does (better) electricity supply increase household enterprise income in India ? 57*, 532–541.

Reardon, T., Berdegué, J., Barret, B. C., & Stamoulis, K. (2007). Household Income Diversification into Rural Nonfarm Activities. In T. Haggblade, Steven; Hazell, P. B. R.; Reardon (Ed.), *Transforming the Rural Nonfarm Economy. Opportunities and Threats in the Developing World* (pp. 115–140). International Food Policy Research Institute.

Regulus. (2020). *Electricity costs in Kenya*. https://stimatracker.com/historic

Renner, M. (2017). *Rural renewable energy investments and their impact on employment*. *1*. http://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_562269.pdf

Roomi, M. S. M., Namal, D. D. A., & Jayasinghe, K. T. (2007). *Study of Energy Consumption Pattern in Sri Lankan Rice Mills - Enhancing Opportunity for Conservation*. *XXXX*(01), 83–88.

Rothman, K. J., Gallacher, J. E. J., & Hatch, E. E. (2013). Why representativeness should be avoided. *International Journal of Epidemiology*, *42*(4), 1012–1014. https://doi.org/10.1093/ije/dys223

Rottger, B. C. & A. (2006). *Small mills in Africa*.

Roy, P., Shimizu, N., & Kimura, T. (2005). Life Cycle Inventory Analysis of Rice Produced by Local Processes. *Journal of JSAM*, *67*(1), 61–67.

Royston, P., & Division, C. (2004). Multiple imputation of missing values. In *The Stata Journal* (Vol. 4, Issue 3).

Sattari, S., Avami, A., & Farahmandpour, B. (2007). *Energy conservation opportunities : Sugar industry in Iran*. 120–126.

Scott, A., Darko, E., Lemma, A., & Rud, J. (2014). *businesses in low and middle income*. *July*.

Shell International. (2017). *SHELL WORLD ENERGY MODEL A VIEW TO 2100*. https://www.shell.com/energy-and-innovation/the-energy-future/scenarios/shell-scenarios-energy-models/world-energy-model/_jcr_content/par/textimage.stream/1510344160326/2ee82a9c68cd84e572c9db09cc43d7ec3e3fafe7/shell-world-energy-model.pdf

Smith, E. G., Janzen, H. H., & Newlands, N. K. (2007). *Energy balances of biodiesel production from soybean and canola in Canada*.

Southall, A. (1980). *Small Urban Centers in Rural Development in Africa*. *9*(9), 237–242.

Spinelli, D., Jez, S., Pogni, R., & Basosi, R. (2013). *Environmental and life cycle analysis of a biodiesel production line from sun fl ower in the Province of Siena (Italy)*. *59*, 492–506.

State Department of Agriculture. (2018). *Early Morning wholesale commodity prices*. Ministry of Agriculture, Livestock and Fisheries.

Szirmai, A., Gebreeyesus, M., Guadagno, F., & Verspagen, B. (2013). *Working Paper Series A review of the literature*. *31*.

Taneja, J. (2018). *If You Build It, Will They Consume? Key Challenges for Universal, Reliable, and Low-Cost Electricity Delivery in Kenya* (CGD Working Paper 491, Issue July 2018).

Terdoo, F., & Feola, G. (2016). *The Vulnerability of Rice Value Chains in Sub-Saharan Africa : A Review*. https://doi.org/10.3390/cli4030047

The World Bank. (2018). *Access to electricity (% of population) - Nigeria*. https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=NG

The World Bank. (2020). *Rural population (% of total population) - Nigeria*. https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=NG

Timilsina, Govinda R; Bhattacharyya, S. C. (2009). *Energy Demand Models for Policy Formulation A Comparative Study of Energy Demand Models*. *March*.

Tran, T., Da, G., Moreno-santander, M. A., Vélez-hernández, G. A., Giraldo-toro, A., & Piyachomkwan, K. (2015). *Resources , Conservation and Recycling A comparison of energy use , water use and carbon footprint of cassava starch production in Thailand , Vietnam and Colombia*. *100*, 31–40.

Ugwuoke, I. C., & Okegbile, O. J. (2014). *Design and Fabrication of Groundnut Shelling and Separating Machine*. *3*(4), 60–66.

UN. (2017). *United Nations Statistics Division - Demographic and Social Statistics. Population density and* *urbaization.* https://unstats.un.org/unsd/demographic/sconcerns/densurb/densurbmethods.htm

UNCTAD. (2019). *ECONOMIC Made in Africa; Economic Development in Africa 2019* (Issue June).

UNICEF Institute for Statistics. (2019). *Sustainable Development Goal 4*. Sustainable Development Goal 4. http://data.uis.unesco.org/

Transforming Our World: The 2030 Agenda for Sustainable Development, A New Era in Global Health (2015). https://doi.org/10.1891/9780826190123.ap02

Urban, B., & Naidoo, R. (2012). Business sustainability: Empirical evidence on operational skills in SMEs in South Africa. *Journal of Small Business and Enterprise Development*, *19*(1), 146–163. https://doi.org/10.1108/14626001211196451

van Ruijven, B. J., Schers, J., & van Vuuren, D. P. (2012). Model-based scenarios for rural electrification in developing countries. *Energy*, *38*(1), 386–397. https://doi.org/10.1016/j.energy.2011.11.037

Wamukonya, N., & Davis, M. (2001). Socio-economic impacts of rural electrification in Namibia: comparisons between grid, solar and unelectrified households. *Energy for Sustainable Development*, *5*(3), 5–13. https://doi.org/10.1016/S0973-0826(08)60272-0

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C. D., Howes, R. E., Tusting, L. S., Kang, S. Y., Cameron, E., Bisanzio, D., … Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, *553*(7688), 333–336. https://doi.org/10.1038/nature25181

White, R. (2002). *"Experience, Strategies, and Project Development." Workshop Synth. Rep., GEF-FAO Workshop on Productive Uses of Renewable Energy.*

Wilcox, M., Waters, L., Wanjiru, H., Pueyo, A., Palit, D., & Rahul Sharma, K. (2015). *Utilising Electricity Access for Poverty Reduction. Annex 2 - Case Study Report: Kenya*. 1–104. http://cdn1.practicalaction.org/e/l/54c7a5a7-3614-4a18-b3c2-16300a0000be.pdf

Williams, N. J., & Jaramillo, P. (2018). Electricity consumption and load profile segmentation analysis for rural microgrid customers in Tanzania. *2018 IEEE PES/IAS PowerAfrica*, 360–365.

Williams, N. J., Jaramillo, P., Taneja, J., & Ustun, T. S. (2015). Enabling private sector investment in microgrid-based rural electrification in developing countries: A review. *Renewable and Sustainable Energy Reviews*, *52*, 1268–1281. https://doi.org/10.1016/j.rser.2015.07.153

World Bank. (2005). *Education sector strategy update. Achieving education for all, broadening our perspective, maximizing our effectiveness*.

World Bank. (2015). Beyond Connections: Energy Access Redefined. *The World Bank*, 1–244. https://mtfenergyaccess.esmap.org/data/files/download-documents/full_report_beyond_connection.pdf%0Ahttp://www.worldbank.org/content/dam/Worldbank/Topics/Energy and Extract/Beyond_Connections_Energy_Access_Redefined_Exec_ESMAP_2015.pdf

Yin, Y., Ma, Z., Nong, G., & Wang, S. (2019). *Strategies of energy management in a cassava starch plant for increasing energy and economic ef fi ciency*. *234*, 1296–1305.

13 Appendix

13.1 Appendix: Relationship between electrification rates and (economic) development

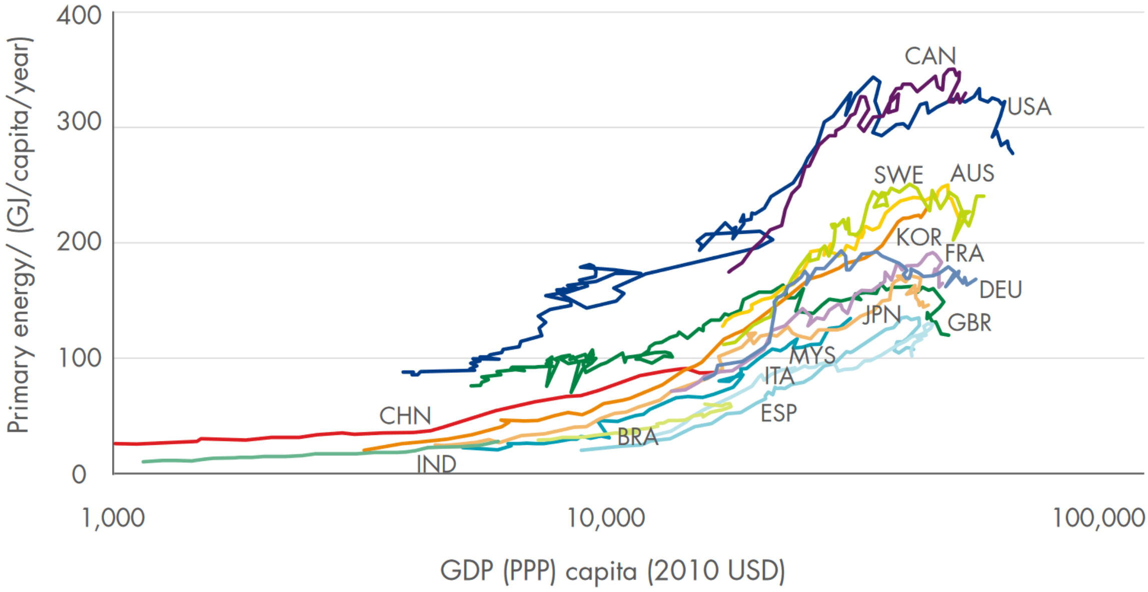

Figure 24. The energy ladder 1960-2016. Source: (Shell International, 2017).

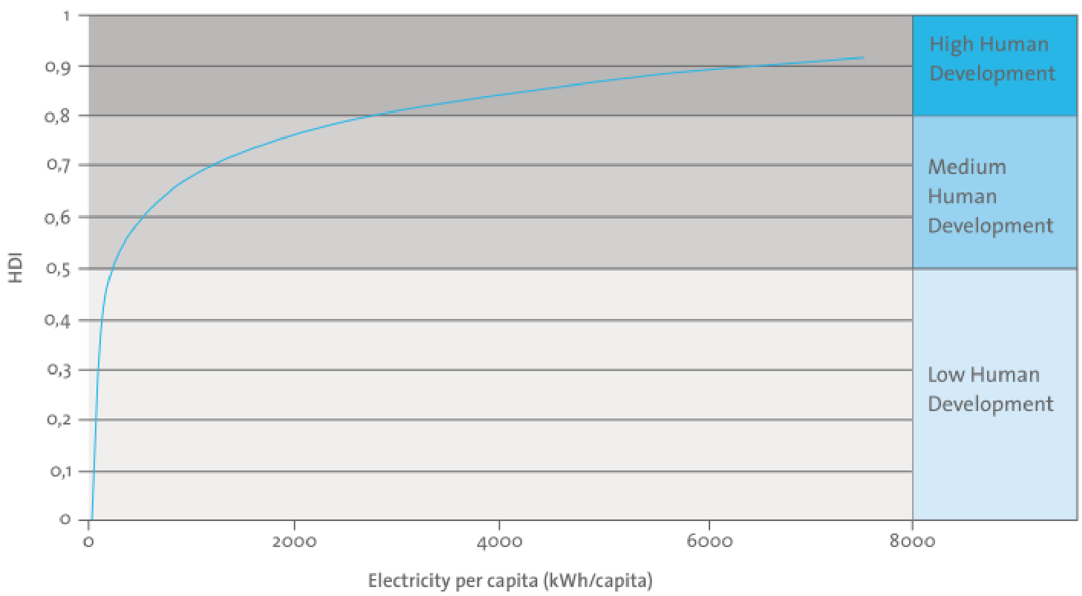

Figure 25. Macro-level correlation between electricity and human development. Source: White (2002).

13.2 Appendix: IMAGE-TIMER and the PrElGen framework

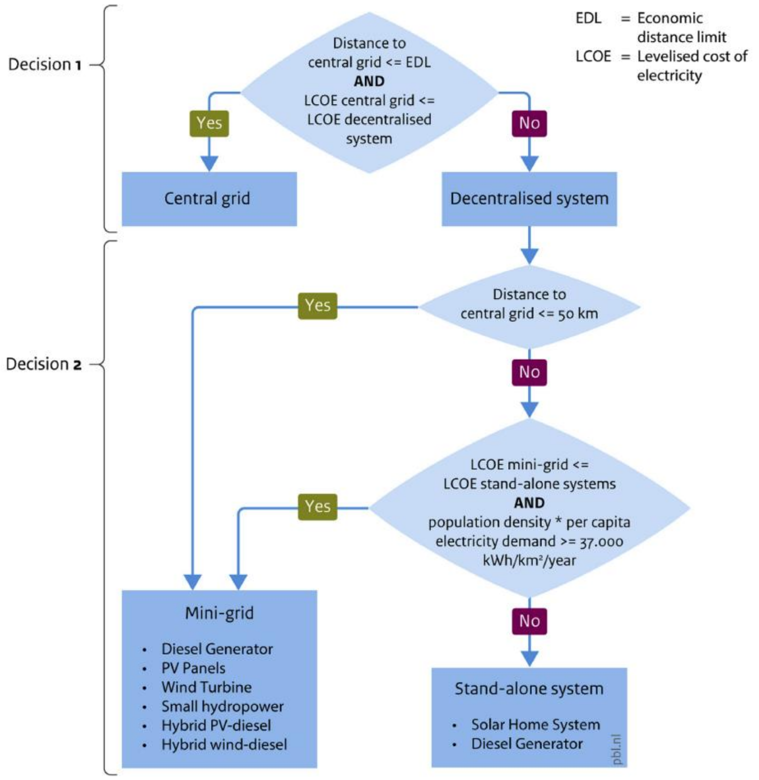

Figure 26. Decision tree to determine the lowest-cost electrification system. Source: Dagnachew et al., 2017.

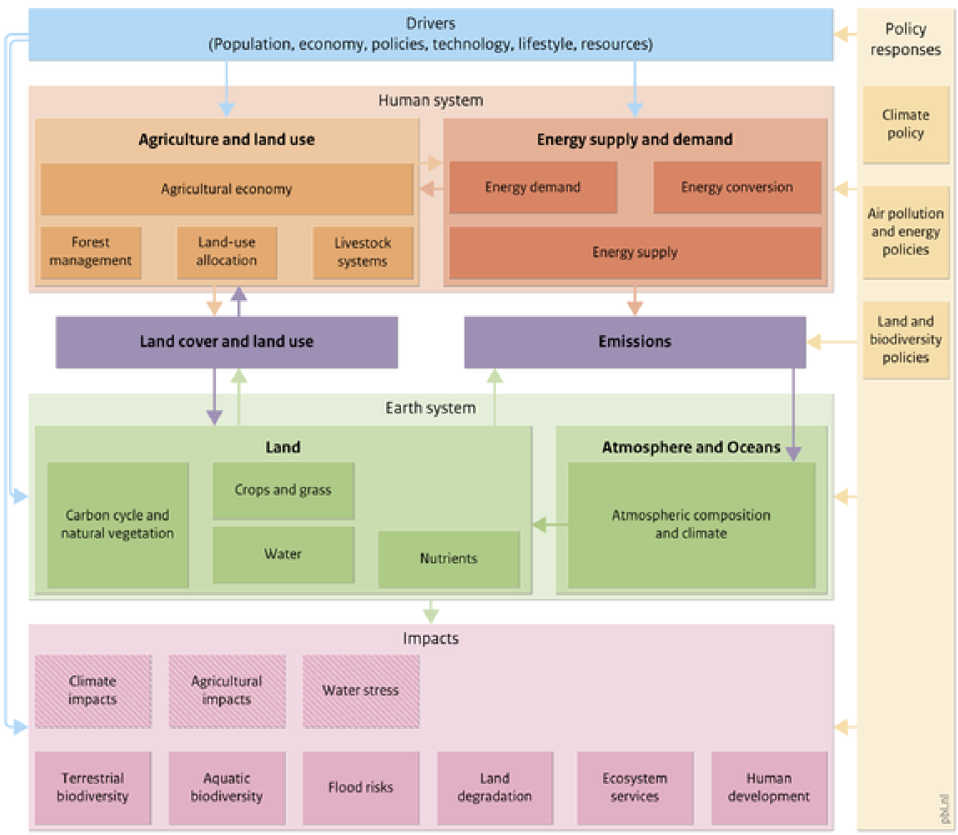

Figure 27. IMAGE 3.0 Framework. Source: PBL, 2014.

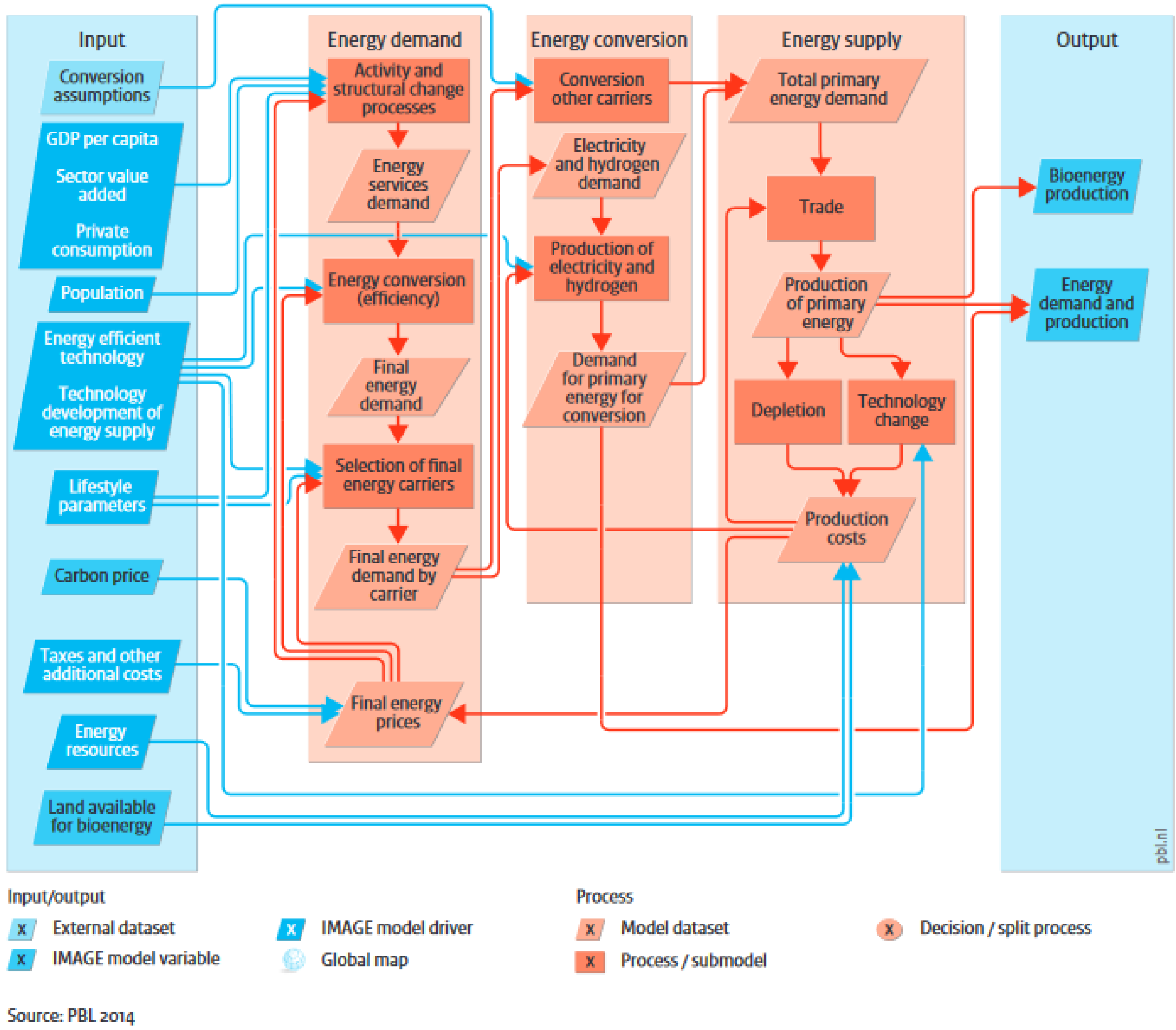

Figure 28. TIMER model. Source: PBL, 2014.

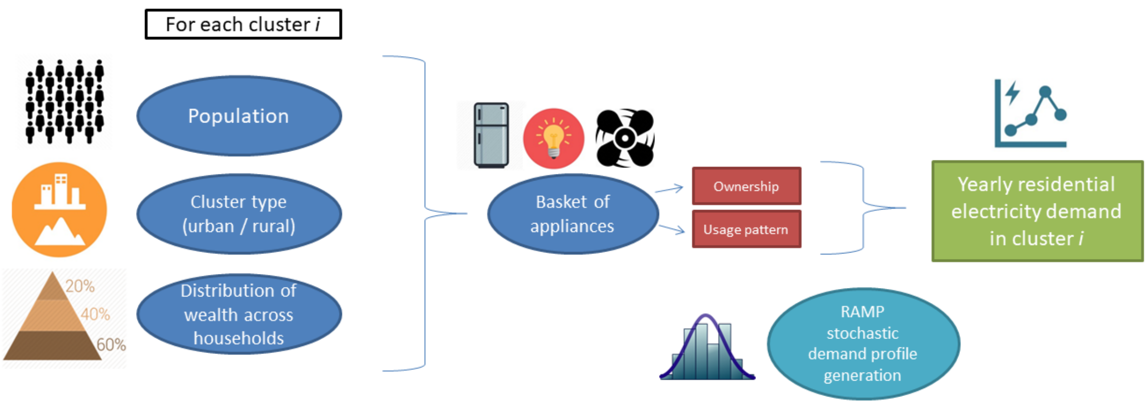

Figure 29. Workflow of the household-type-differential residential electricity requirement estimation procedure. Source: (Falchetta, n.d.).

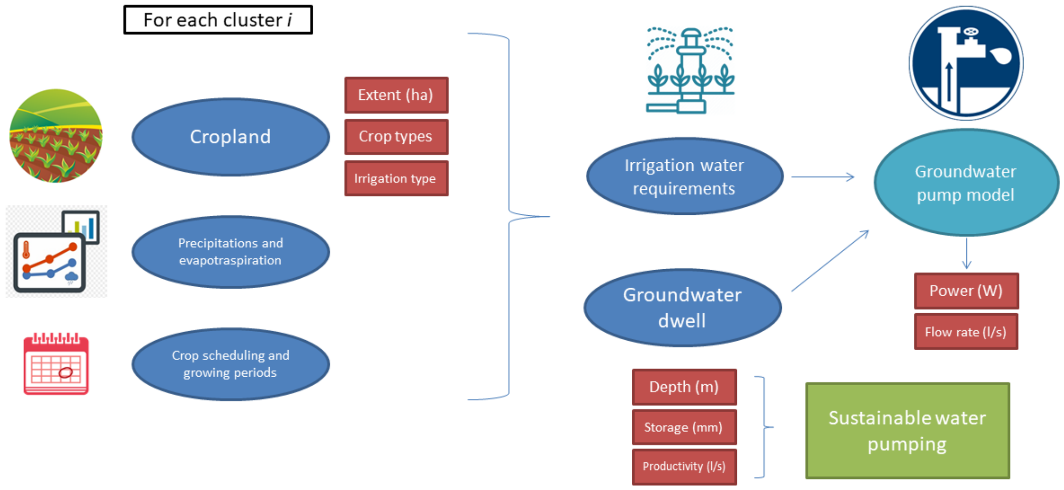

Figure 30. Workflow of water pumping electricity requirement estimation procedure. Source: (Falchetta, n.d.).

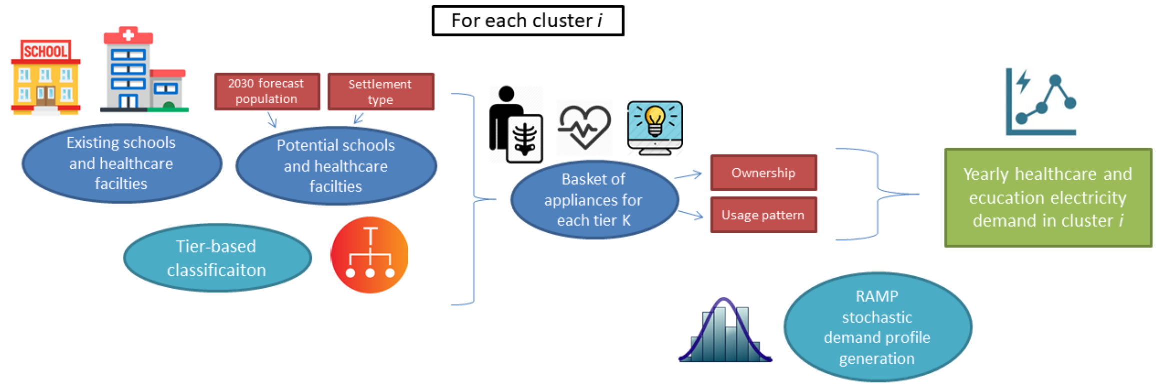

Figure 31. Workflow of the tier-differentiated healthcare and education electricity requirement estimation procedure. Source: (Falchetta, n.d.).

13.2.1 Assumptions of the Shared Socioeconomic Pathways Scenarios

The Shared Socioeconomic Pathway 2 (SSP2) scenario represents moderate assumptions about demographic and socioeconomic parameters such as population growth, economic growth and acceptance of energy technologies. The different SSPs are explained in O'Neill et al. (2014) and contain not only information on socio-economic factors but also their consequences on GHG emissions. The scenarios following different storylines concerning two dimensions: challenges for GHG mitigation and challenges for adaption. SSP1 assumes little challenges about both adaption and mitigation as consequence of different interacting factors such as fast income growth, reduced reliance on natural resources, quick technological change, high levels of international cooperation, high levels of education, low fertility rates and thus smaller populations. SSP2, as previously stated assumes a moderate development with little changes from the current situation. SSP3 assumes slow economic development due to lesser degrees of the previously stated factors and severe mitigation as well as adaption challenges. SSP4 assumed higher levels of growth in high-income countries, thus better capacities to mitigate, and lower levels of growth in developing countries, thus greater challenges to adapt. SSP5 assumes full focus on economic development but no focus on the environmental consequences, which leads to large challenges for mitigation but relatively low adaption challenges. The impact of differing assumptions is supposed to be tested by running a scenario analysis with SSP1 and SSP3 scenarios.

13.3 Appendix: Crop Processing

The format, detail and depth of data provided was quite varied across the different papers. The following section therefore explains in more detail how studies were selected and how missing information was dealt with. Ideally, the studies would have been focused on only micro-scale processing facilities in SSA or at least developing countries and represent the most important processed goods according to their consumption in different regions. However, little quantitative data is available for these conditions, so compromises had to be made, which are outlined in the following section.

Milling off different grains, legumes and cassava can lead to different degrees of coarseness, e.g. for wheat can be milled to flour but also the coarser semolina or bran. Generally, finer flour is preferred in East Africa, while people in Central and West Africa tend to consume more coarse products (Rottger, 2006). Following the availability of data, the final product for wheat and maize assumed in this study is flour, except for one facility which also produced some semolina and bran from wheat. For cassava, the milled product is assumed to be starch as most common product. Millet is also assumed to be milled but to a coarser form than the other crops. Rice is usually consumed either raw or parboiled (Terdoo & Feola, 2016), which are therefore products this study is focusing on.

Sugar beet, as drought-resistant crop is gaining increasing attention especially in countries such as Kenya (Mandere et al., 2011), where there is, in fact a shortage in sugar crops and a realistic potential to substitute imports with local production. This also favors the cultivation of sugarcane, which thrives most in high exposure to sunlight (Pippo & Luengo, 2016). Next to producing sugar, there are some other applications, e.g. sugar beet can also be used unprocessed as animal fodder and sugar cane can be used to make sugar cane juice. However, given the relevance of sugar production and the lack of data on other products and their processes, the focus of this paper is on the production of sugar.

The only product from rapeseed considered here is oil. Even though some studies also give information on the transformation of oil to biodiesel, this research only focuses on the production of food oils, as the other stories only refer to developed countries. A similar situation holds for sunflowers. They can also be consumed unprocessed, but the typical form of consumption is as oil.

There is a vast range of products that can be made from soybeans. In 2004, about 2/3 of imported soybean products were oil and 1/3 were grain/meal (Lawrence, 2011). According to this and the limited data availability, the products from soybeans in this study are oil and soybean meal.

Ground nuts, field peas and other pulses are mostly eaten whole but they are also used as meal (Kuwonu, 2019), therefore most selected sources just focus on dehulled, dried and split peas and ground nuts and one source refers to the process for ground peas. Groundnuts can also be processed into oil and meal but due to lack of data, these options are not included (Nautiyal, 2002).

Further uncertainty is added because processing for the same product can be done by different operations and different operations can be electrified to different degrees. Unfortunately, no literature was found which would have presented an encompassing overview of possible operations so again, it had to be decided by using common sense, if the processes described by different sources seemed feasible for a micro-scale farmer in SSA. Therefore, it is also impossible in the scope of this study to say anything about the conditions under which it would actually be profitable to use electrical energy for processing but only about the potential.

Most used sources were energy audits of processing sites, some of them introduced improved machinery to increase energy efficiency. Both kinds of articles were included in the review. Many studies only provided information on the overall energy of a facility. However, since these also include human labor and thermal energy, not just electricity, those studies were excluded. Often, some key information was missing on either the selection criteria (namely spatial scope and facility size) or electricity consumption per unit input. If information on only either country scope or size was missing or not matching the criteria, studies were included anyway. The same was the case if the scale was not explicitly given but could be inferred from other information. However, if the scale was too large and it seemed like the required machinery was too expensive and specific, the studies were excluded. If information on electricity consumption was missing but it could be compensated with some simple assumptions, the studies were included. Ideally, only the most recent studies would have been selected, however, this would have limited the available data too much. About half the studies were published after 2010, the rest between 2010 and two studies were published in 1984 and 1989. However, simple mechanical operations e.g. of milling machines do not seem to have changed much when comparing with more current data.

A general simplification that was made, if a range of values was given (e.g. production output between 100 and 110 kg/h of milled millet according to (Balasubramanian, 2015), the average was taken for the calculation of energy consumption, i.e. 105 kg/hr.

Since the size of the facility was not always indicated directly, it as approximated by using available data and where data was not available, assumptions were made. The final criteria for size was sales < 100,000 USD/year. If not directly indicated, this was calculated based on the possible output times the expected product.

$$annual\ sales\ (USD/year)\ =\ annual\ output\ (kg/year)\ *\ price\ of\ product\ (USD/kg)$$

For prices of products, data from Kenya was used (State Department of Agriculture, 2018). If the total annual output was not given, it was calculated based on the output/unit of time given, usually hours, and multiplied to give the output per year based on a 6 hours workday for 300 days/year, i.e. 1800 hours of operation.

$$annual\ output\ (kg/year)\ =\ output\ per\ hour\ (kg/h)\ *\ hours\ per\ year\ (h/year)$$

This is a relatively strict assumption since due to seasonality of crop harvests, the actual number of days per year on which the crops can be processed is probably lower so facilities with a larger output per hour might still be below the 100,000 USD threshold. However, since only some of the studies actually addressed facilities of this order of magnitude, no paper was actually excluded because of this criterion. The differentiation rather serves to get transparency about the impact the scale of a facility might have on its electricity consumption.

Further assumptions had to be made to calculate the actual electricity consumption per unit of crop input. If electricity consumption was not given explicitly, the electricity requirements for a processing step were calculated based on the rate power of an engine, which was given in hp (horsepower) or W. (e.g. a 1kW engine) and the expected hours of operations, which were again 1800 hours/year. Some studies did not indicate the input but rather the throughput or output. In these cases, if provided, the production efficiency/recovery rate per step (i) was used to calculate the input. If a study did not provide the production efficiency itself, it was assumed the be equal to that given in other studies on the same crop and product.

$$producitonefficiency_i = \frac{output_i}{input_i}$$

The total production efficiency of a production process is the product of the efficiencies of the single operations.

$$productionefficiency_{total} = \sum_{n=i} productionefficiency_i$$

Sometimes the efficiency was not given for all processing steps, in which case they were assumed to be the same for all processing steps. Sources furthermore provided information of capacity (input, throughput or output) for different units of time and in some cases per run of the facility rather than per unit of time. If given per unit of time, this had to be matched with the electricity consumption in the same unit of time. As for the calculation of scale, it was sometimes necessary to make assumptions about the hours of operation, for which again, 1800 hours/year were assumed. If a source indicated the capacity per run, sometimes there was also information on the length of one run, so the same calculation could be done as if capacity was given per time. If this information was lacking, the source was not used.

Table 13. Average, maximum and minimum electricity consumptions for different crops compared with data form the World Bank.

	Electrical Energy consumption in MJ/kg			**According to Banerjee et al. (2017)**	**Divergence of mean from Banerjee et al. (2017)**
	AVG	**Max**	**Min**		
Groundnut	0.032	0.050	0.007		
Field peas	0.039	0.083	0.018		
Rice	0.052	0.128	0.011	0.18	0.290
Cassava	0.064	0.166	0.011	0.054	1.177
Wheat	0.076	0.104	0.023	0.432	0.175
Millet	0.104	0.233	0.012		
Sunflower	0.120	0.193	0.060	0.144	0.836
Rapeseed	0.179	0.296	0.023	0.144	1.242
Maize	0.184	0.491	0.029	0.216	0.854
Soybean	0.210	0.319	0.101	0.144	1.457
Sugar beet	0.249	0.475	0.083		
Sugar cane	0.334	0.594	0.090	0.144	2.323

Table 14. Crop processing findings with indication where own assumptions of the author were used for the assessment (a).

Source	**Crop**	**Country/Area**	**Size**	**Product**	**Electricity demand (MJ/kg)**
(Olaoye et al., 2020)	Wheat	Nigeria	Micro	Flour (75%), semolina (3%), bran 22%)	0.101
(Green et al., 2019)	Wheat	Nigeria	Medium/Large (a)	Flour	0.023
(Bassey & Schmidt, 1989)	Wheat	South Africa	Small (a)	Flour	0.104 (a)
(Goyal et al., 2010)	Rice	India	Medium/Large (a)	Raw milled rice, parboiled rice	0.057 (a)

Source	Crop	Country/Area	Size	Product	Electricity demand (MJ/kg)
(Goyal & Agrawal, 2008)	Rice	India	Medium/Large (a)	Raw milled rice	0.019
(Roy et al., 2005)	Rice	India	Medium (a)	Raw milled rice, parboiled rice	0.076
(Ahiduzzaman & Islam, 2009)	Rice	Bangladesh	Medium/Large	Parboiled rice	0.025 (a)
(Roomi et al., 2007)	Rice	Sri Lanka	Medium (a)	Raw milled rice, parboiled rice	0.087 (a)
(Bassey & Schmidt, 1989)	Rice	South Africa	Micro/Small	Raw milled rice	0.052
(ILO et al., 1984)	Maize	Developing Countries	Micro/Small	Maize flour	0.031 (a)
(Rottger, 2006)	Maize	Africa	Micro	Maize flour	0.079 (a)
(Abubakar et al., 2010)	Maize	Nigeria	Micro	Maize flour	0.491 (a)
(Aliu et al., 2018)	Maize	Nigeria	Large	Maize flour	0.231
(Dozier et al., 2005)	Millet	US	Medium/Large	Milled pear millet	0.012 (a)
(Ambrose et al., 2017)am	Millet	India	Micro	Flour from different millet types	0.201 (a)
(Balasubramanian, 2015)	Millet	India	Micro	Milled millet	0.024
(Chin et al., 2012)c	Millet	India	Micro/Small	Milled millet	0.026
(Bassey & Schmidt, 1989)	Millet	South Africa	Small/Medium	Milled millet	0.090
(Opoku & Tabil, 2006)	Field Peas	Canada	Micro	Ground field peas	0.045 (a)
(Lal & Verma, 2007)	Field Peas	India	Micro/Small	Dried, split pulses	0.030 (a)
(Goyal & Agrawal, 2008)	Field Peas	India	Micro/Small	Dried, split pulses	0.032 (a)
(Bassey & Schmidt, 1989)	Field Peas	South Africa	Micro/Small	Dehulled cowpea	0.083
(Chouhan & Chandrakar, 2014)	Sugar Beet	India	Large (a)	Sugar	0.101 (a)
(Jensen et al., 2015)	Sugar Beet	Europe	Large (a)	Sugar	0.083 (a)
(Sattari et al., 2007)	Sugar Beet	Iran and world	Medium (a)	Sugar	0.407 (a)
(Jekayinfa & Olajide, 2007)	Cassava	Nigeria	Small (a)	Starch	0.011
(Tran et al., 2015)	Cassava	Thailand, Vietnam, Colombia	Small/Medium	Starch	0.047 (a)
(Da et al., 2013)	Cassava	Colombia and Vietnam	Medium (a)	Starch	0.030
(Yin et al., 2019)	Cassava	Africa and Asia	Large	Starch	0.166 (a)
(Spinelli et al., 2013)	Sunflower	Italy	Large (a)	Oil	0.193
(Ion & Popescu, 2017)	Sunflower	unclear	Large	Oil	0.060
(Kartika et al., 2010)	Sunflower	France	Large	Oil	0.108
(Bassey & Schmidt, 1989)	Soybean	South Africa	Small	Milled soybean	0.101
(Smith et al., 2007)	Soybean	Canada	Large (a)	Oil	0.319
(Eshetu et al., 2017)	Groundnuts	Ethiopia	Small (a)	Shelled groundnuts	0.040
(Ugwuoke & Okegbile, 2014)	Groundnuts	Nigeria	Small	Shelled groundnuts	0.007
(Atiku et al., 2004)	Groundnuts	Nigeria	Small	Shelled groundnuts	0.050
(Smith et al., 2007)	Rapeseed	Canada	Large (a)	Oil	0.217

Source	Crop	Country/Area	Size	Product	Electricity demand (MJ/kg)
(Fore et al., 2011)	Rapeseed	US	Medium (a)	Oil	0.023
(Chen & Chen, 2011)	Rapeseed	China	Large (a)	Oil	0.296 (a)
(M. Soppimath & G. Hudedmani, 2017)	Sugar Cane	India	Large	Sugar	0.230
(Pippo & Luengo, 2016)	Sugar Cane	unclear	Large	Sugar	0.090
(Sattari et al., 2007)	Sugar Cane	Iran and world	Medium (a)	Sugar	0.509 (a)

13.4 Appendix: Literature Review

Table 15. Predictors with sources and indication of relevant dependent variables.

Category	Predictor	Source
Household Characteristics		
	Income/level of wealth/assets	ADA (2016) (U)
		Nagler and Naudé (2018) (U)
		Christiaensen and Demery (2018) (U)
		Kooijman-van Dijk and Clancy (2010) (C)
		Nkegbe et al. (2018) (U)
	number of hh members	Owoo and Naudé (2015) (P)
		Nagler and Naudé, (2017) (U)
		Christiaensen and Demery (2018) (U)
	number of rooms	Nagler and Naudé, (2017) (U)
	shock experience	Efobi et al. (2019) (U)
		Nagler and Naudé, (2017) (U)
		Christiaensen and Demery (2018) (U)
	house size	Rao (2013) (U)
	agricultural activity (precipitation)	Owoo and Naudé (2015) (P)
		Nagler and Naudé (2017) (U)
		Christiaensen and Demery (2018) (U)
	access to information (e.g. radio ownership)	Nkegbe et al. (2018) (P)
Entrepreneur Characteristics		
	knowledge/skills (general)	Practical Action Consulting, (2012) (P, C)
		Wilcox et al. (2015) (P, C)
		Cabraal et al. (2005) (P, C)
		Naidoo and Urban, (2012) (P)
		Pueyo et al. (2014) (C)
		Mayer-Tasch et al. (2013) (P)
	level of education (years of schooling, literacy)	Rao (2013)
		Mayer-Tasch et al. (2013) (P)
		ADA (2016) (U)
		Efobi et al. (2019) (U)
		Nkegbe et al. (2018) (U)
		Christiaensen and Demery (2018) (U)
		Nagler and Naudé, (2017) (U)
		Haggblade (2007) (U)
	digital and mechanical know-how	Mayer-Tasch et al. (2013) (P)
		Blodgett (2016) (U, P, C)
	social networks (religion)	Kooijman-van Dijk (2012) (P)
		Owoo and Naudé (2015) (U, P)
		Kooijnman-van Dijk (U)
		Nkegbe et al. (2018) (U)
		Practical Action Consulting, (2012) (P)
	marital status	ADA (2016) (U)
		Efobi et al. (2019) (U)
		Akpan et al. (2014) (U)
		Nagler and Naudé, (2017) (U)
	age	Efobi et al. (2019) (U)
		Owoo and Naudé (2015) (P)
		Christiaensen and Demery (2018) (U)
		Akpan et al. (2014) (U)
		Blodgett (2016) (U, P, C)

Category	Predictor	Source
	sex	Lecoque & Wieman (2015) (C)
		(Maleko, 2005)Maleko (2005) (P)
		Owoo and Naudé (2015) (P)
		Nagler and Naudé, (2017) (U, P)
		Christiaensen and Demery (2018) (U)
		Mapako and Prasad (2008) (U)
		Harsdorff and Bamanyaki, (2009) (C)
		Akpan et al. (2014) (U)
		Mead and Liedholm (1998) (U)
	awareness and acceptance of productive uses of electricity	Renner (2017) (C)
		Mayer-Tasch et al. (2013) (C)
	personal character	ADA (2016) (U, P)
	level of entrepreneurial experience	Benedikter et al. (2013) (U)
		ADA (2016) (U)
		Lecoque & Wieman (2015) (C)
Enterprise Characteristics		
	Wealth, size (income, assets, capital)	Mayer-Tasch et al. (2013) (C)
		Kooijman-van Dijk and Clancy (2010) (C)
		Mapako and Prasad (2008) (P)
		Harsdorff and Bamanyaki (2009) (C)
		Ekblom (2016) (P)
		Pueyo et al. (2014) (C)
		Banerjee et al. (2017) (C)
		Mead and Liedholm (1998) (P)
	type/industry	Mayer-Tasch et al. (2013) (C)
		Kooijnman-van Dijk (2008) (P, C)
		Owoo and Naudé (2015) (P)
		Mead and Liedenholm (1998) (P)
		Pueyo et al. (2014) (C)
		Harsdorff and Bamanyaki, (2009) (C)
	registration and formality	ADA (2016) (P)
		Loening and Lane (2007) (P)
		Foster et al., (2010) (P)
	location of enterprise	Olawale and Garwe (2010) (P)
		Mayer-Tasch et al. (2013) (P)
		Nagler and Naudé, (2017) (U)
		Christiaensen and Demery (2018) (U)
		Pueyo et al. (2014) (C)
		Blodgett (2016) (U, P, C)
Market Access: Demand		Kooijman-van Dijk and Clancy (2010) (P, C)
		Harsdorff and Bamanyaki, (2009) (P)
		Mead and Liedenholm, (1998) (U)
	population density	Kooijman-van Dijk and Clancy (2010) (P, C)
		Kooijman-van Dijk and Clancy (2010) (C)
		Haggbladde (200/) (U)
	urban/rural migration	Nagler and Naudé, (2017) (U)
	local wealth and purchasing power	Practical Action Consulting, (2012) (P)
	local population/market size	Mayer-Tasch et al. (2013) (P)
	number of people who transit	Blodgett (2016) (U, P, C)
	level of competition	Practical Action Consulting, (2012) (P)
Market Access: Supply/Input factors		Practical Action Consulting, (2012) (P)
	access to finance (presence of banks, interest rate, required collateral, required owners' equity contribution)	Mayer-Tasch et al. (2013) (P)
		ADA (2016) (U)
		FRES (2014) (U)

Category	Predictor	Source
		Nagler and Naudé, (2017) (U)
		Christiaensen and Demery (2018) (U)
		Blodgett (2016) (U)
		Fishbein et al. (2003) (U)
		Ekblom (2016) (P)
		Pueyo et al. (2014) (C)
		Loening and Lane (2007) (P)
		Olawaleand Garwe (2010) (U, P)
		Mead and Liedenholm (1998) (U)
	cost of input factors (other than electricity)	Pueyo et al. (2014) (C)
		Blodgett (2016) (C)
	presence electricity access	Rao (2013) (P, C)
		Mayer-Tasch et al. (2013) (P)
		Maleko (2005) (U)
		Prasad and Dieden (2007) (U,P)
		FRES (2013) (U)
		Harsdorff and Bamanyaki, (2009) (U)
		Blodgett (2016) (U)
		Peter et al., (2011) (U)
		Kirubi et al., (2008) (P)
		Lenz (2017) (U)
		Gibson and Olivia, (2015) (U)
	electricity access capacity (voltage, AC/DC…) quality (number and duration of blackouts in given time; presence of risk factors for supply)	Mayer-Tasch et al. (2013) (P)
		World Bank (2015) (C)
		Foster et al., (2010) (P)
		Blodgett (2016) (C)
		Mapako and Prasad (2008) (P)
		Moyo (2013) (P)
		Pueyo et al. (2014) (C)
		World Bank (2015) (C, P)
		Gibson & Olivia, 2010) (U)
	cost of electrical connection	Mayer-Tasch et al. (2013) (C)
		Pueyo et al. (2014) (C)
	cost of electricity consumption	Pueyo et al. (2014) (C)
Physical Market Access and Transportation		Rao (2013) (P, C)
		Wilcox et al. (2015)
		Mayer-Tasch et al. (2013) (P)
		Nkegbe et al. (2018) (U)
		Haggblade (U)
	physical access to input goods (e.g. appliances) and services (e.g. maintenance)	MFAN, (2014) (P)
		Mayer-Tasch et al. (2013) (P, C)
		Peter et al., (2011) (P, C)
	level of urbanization	Rao (2013) (P, C)
		Mead and Liedenholm (1998) (P)
		Akpan et al. (2014) (U)
	distance to urban centers	Nagler and Naudé, (2017) (U)
		Christiaensen and Demery (2018) (U)
		Gibson and Olivia, (2015) (U)
	presence of roads and distance to roads	Kooijman-van Dijk and Clancy (2010) (C)
		Kooijman-van Dijk and Clancy (2010) (C)
		Loening and Lane (2007) (P)
		Efobi et al. (2019) (U)
	quality of roads	Kooijman-van Dijk (2012) (C, P)
		Mayer-Tasch et al. (2013) (P)
		Loening and Lane (2007) (P)

Category	Predictor	Source
		Gibson and Olivia, (2015) (U)
	presence of regular market/commercial center	FRES (2014) (U)
		Mayer-Tasch et al. (2013) (P)
	access to daily public transportation	diaoDiao et al. (2018) (U)
Governance, institutional context and development		ADA (2016) (U)
	conductive regulatory environment and policies	Practical Action Consulting, (2012) (P)
		ADA (2016) (U)
		Lecoque & Wieman (2015) (C)
		Nagler and Naudé (2018) (P)
		Olawaleand Garwe (2010) (U, P)
		Blodgett (2016) (U)
	effort to start a business	World Bank (2018) (U, P)
	level of corruption	Olawaleand Garwe (2010) (U, P)
		ADA (2016) (U)
	political relevance and activity of locality (presence of communal administration)	Mayer-Tasch et al. (2013) (P)
	level of social security	Christiaensen and Demery (2018) (U)
	availability of subsidies to SMEs	ADA (2016) (U)
	country	Owoo and Naudé (2015) (P)
	GDP	Chu et al. (2007) (U, P)
	gender equality	Etim and Iwu (2018) (U)
	presence of violent conflicts	Loening and Lane (2007) (P)
		Nagler and Naudé (2018) (P)
	access to water	Cabraal et al., 2005 (P)
	presence of schools	Cabraal et al. (2005) (P)
		Mayer-Tasch et al. (2013) (P)
	presence of health facilities	Cabraal et al. (2005) (P)
	investment climate	Aterido (2010) (U)
	access to cell phone communication	Loening and Lane (2007) (P)
Other		
	season	Loening and Lane (2007) (P)
		Blodgett (2016) (U)
	weather (droughts)	Blodgett (2016) (C)
Development Program Characteristics		Aterido (2013) (P)
	training and information provided and quality (BDS), knowledge sharing	Mayer-Tasch et al. (2013) (P, C)
		ADA (2016) (U)
		Lecoque & Wieman (2015) (C)
		Tanzania National Survey Baseline report (2012) (P)
		Blodgett (2016) (U)
		Fishbein (2003) (U)
	monitoring progress and satisfaction	Golumbeanu and Barnes (2014) (P, C)

13.5 Appendix: Additional Information on Data Sources

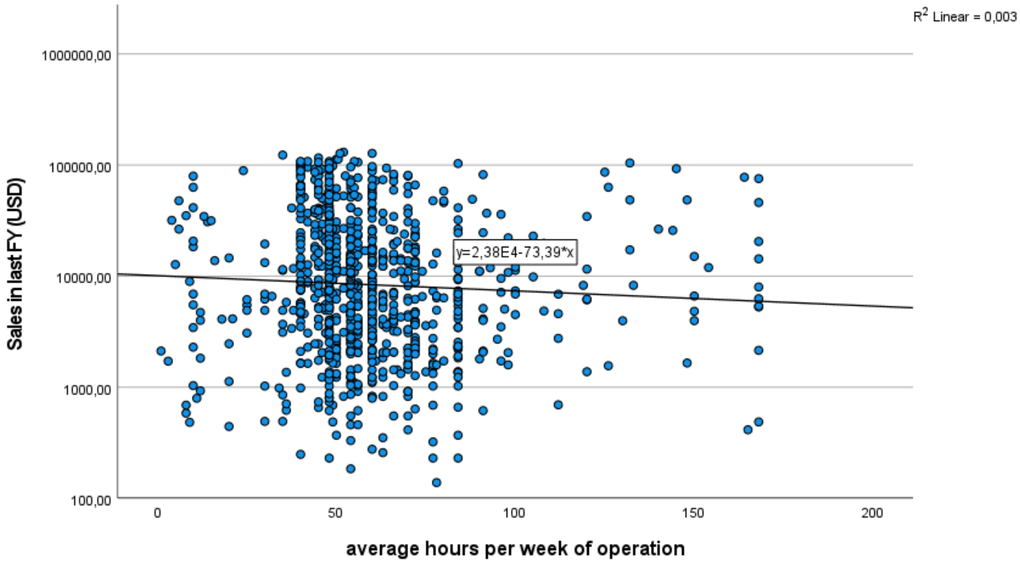

Figure 32. Correlation between total annual sales and average hours of operation per week

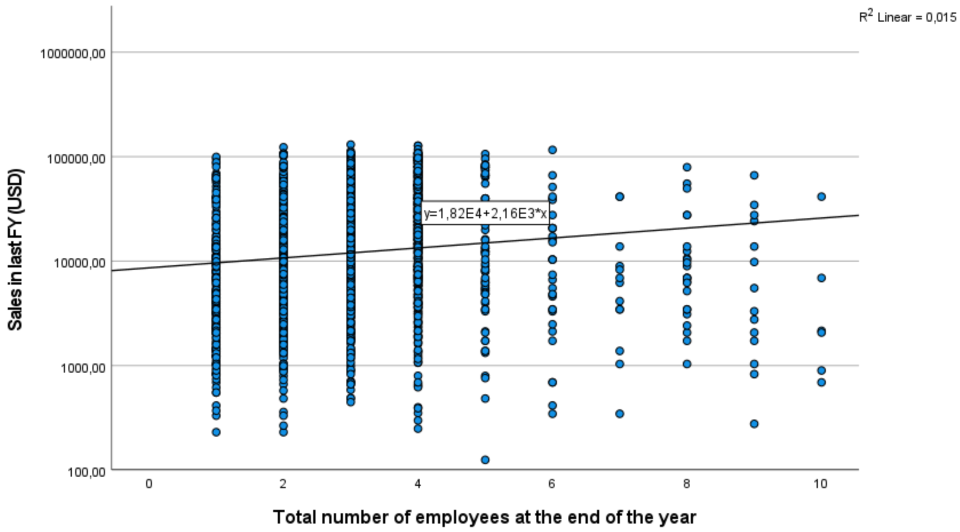

Figure 33. Correlation between total annual sales and the total number of full-time employees at the end of the year.

Table 16. Electricity tariffs and exchange rates with sources.

Country	Electricity price LCU/kWh	LCU	Year	Electricity price USD/kWh 2020	Exchange rate LCU to USD	Source
Burkina Faso	88.00	West African CFA franc	2010	0.23	448.35	(C. Briceño-Garmendia & Domínguez-Torres, 2010)
Cameroon	83	Central African CFA franc	2008	0.22	448.35	(ENEO, 2015)
DRC	110.40	Congolese Franc	2013	0.22	553.98	(Huenteler et al., 2017)
Eswatini	1.45**	Lilangeni	2006	0.30	6.38	(ESI Africa, 2019) (EEC, 2018)
Gambia	7.76***	Dalasi	2006	0.36	28.5	(Ceesay, 2012; NAWEC, 2020; PURA, 2008)
Kenya	19.12	Kenyan Shilling	2012	0.25	84.43	(Regulus, 2020)
Mozambique	4.35	Metical	2019	6.2	63.43	(GlobalPetrolPrices.com, 2020)
Namibia	0.40	Rand	2005	0.08	6.38	(ECA, 2009; LAC, 2008)
Nigeria	-	-	-	-	159.29	
Kano	22.84	Naira	2013	0.15	159.29	(Okafor, 2016)
Abia	21.78	Naira	2013	0.15	159.29	(Okafor, 2016)
Abuja	21.03	Naira	2013	0.14	159.29	(Okafor, 2016)
Kaduna	23.85	Naira	2013	0.15	159.29	(Okafor, 2016)
Lagos	20.91	Naira	2013	0.15	159.29	(Okafor, 2016)
Cross river	22.4	Naira	2013	0.16	159.29	(Okafor, 2016)
Anambra	20.11	Naira	2013	0.16	159.29	(Okafor, 2016)
Oyo	22.84	Naira	2013	0.14	159.29	(Okafor, 2016)
Enugu	21.78	Naira	2013	0.16	159.29	(Okafor, 2016)
Katsina	22.84	Naira	2013	0.16	159.29	(Okafor, 2016)
Jigawa	22.84	Naira	2013	0.16	159.29	(Okafor, 2016)
Kebbi	23.85	Naira	2013	0.16	159.29	(Okafor, 2016)
Kwara	22.84	Naira	2013	0.16	159.29	(Okafor, 2016)
Niger	21.03	Naira	2013	0.16	159.29	(Okafor, 2016)
Ogun	22.84	Naira	2013	0.15	159.29	(Okafor, 2016)
Zamfara	23.85	Naira	2013	0.15	159.29	(Okafor, 2016)
Nasarawa	21.03	Naira	2013	0.16	159.29	(Okafor, 2016)
Sokoto	23.85	Naira	2013	0.16	159.29	(Okafor, 2016)
Gombe	22.84	Naira	2013	0.16	159.29	(Okafor, 2016)
Rwanda	113.46	Rwanda Franc	2012	0.23	582.15	(Mawejje et al., 2012)
Uganda	414	Ugandan Shilling	2016	0.31	1774.9	(Mawejje et al., 2012)
Zimbabwe*	0.10	Zimbabwe Dollar	2015	0.11	1	(Bungane, 2016)
Cote d'Ivoire	66.96	West African CFA franc	2009	0.27	448.35	(Abidjan.net, 2020)
Cape Verde	26.4f	Cape Verde Escudo	2009	0.33	75.36	(C. M. Briceño-Garmendia & Benitez, 2011)

Source for exchange rates: https://fxtop.com/en/currency-converter.php, source for inflation rates
If possible, tariffs for small businesses with medium consumption levels were applied
* Zimbabwe was undergoing a change in currency and the WB could not say for sure which LCU was used in the surveys
** For Eswatini, linear growth between two years was assumed and extrapolated to 2006
*** As tariff seems stable over time, tariff of 2008 was assumed and 2006 exchange rate

Table 17. Years of DHS used in the research per country.

Country	DHS Survey
Burkina Faso	2010 DHS
Cameroon	2018 DHS
Congo Democratic Republic	2013-14 DHS
Eswatini	2006-07 DHS
Gambia	2013 DHS
Kenya	2014 DHS
Mozambique	2011 DHS
Namibia	2013 DHS
Nigeria	2018 DHS
Rwanda	2014-15 DHS
Uganda	2016 DHS
Zimbabwe	2015 DHS
Cote d'Ivoire	2011-12 DHS

Table 18. All variables found with sources and feasibility of modelling.

Category	Predictor	Variables found	Source	Modelling possible?	Level	Modelling possible per regression			
						U	Conn	S	Cons
Household characteristics									
	Income/level of wealth/assets	share of population by wealth quintile	DHS	yes	state	y	y	y	y
		value of dwelling and different assets	WB HH survey	no, no data for SSA	HH	n	n	n	n
		wealth index gini coefficient	DHS	yes	state	y	y	y	y
		income from different sources	WB HH survey	no, no data for SSA	HH	n	n	n	n
		value and possession of different assets	WB HH survey	no, no data for SSA	HH	n	n	n	n
		share of population who possess different assets	DHS	no, no sufficient represent ations	state	n	n	n	n
	number of hh members	number of HH members	WB HH survey	yes	HH	y	y	n	n
		share of population per household size	DHS	yes	state	y	y	-	-
	number of rooms	number of rooms	WB HH survey	no, no data for SSA	HH	n	n	n	n
	shock experience	dichotomous variables on the occurrence of different shocks on community and hh level	WB HH survey	yes	HH	y	y	n	n
		occurrence of droughts	Beguería and Serrano (2016)	yes	grid-cell	y	y	n	n
		occurrence of floods	Beguería and Serrano (2016)	yes	grid-cell	y	y	n	n
	house size	-	-	-	-	-	-	-	-
	agricultural activity (precipitation)	agricultural activity of any household member (dichotomous)	WB HH survey	yes	HH	y	y	n	n
		share of population occupied in agriculture (m/f)	DHS	yes	state	y	y	y	y
	access to information (e.g. radio ownership)	share of HH who own a radio	DHS	yes	state	y	y	y	y
		indication if a HH owns a radio	WB HH Survey	yes	HH	y	y	y	y
Entrepreneur characteristics									
	knowledge/skills (general)	-	-	-	-	-	-	-	-
	level of education (years of schooling, literacy)	attendance of any school (dichotomous)	WB HH survey	yes	individual	y	y	n	n

		Highest level of education of top manager	WB EP Survey	no, too many missing values	EP	n	n	n	n
		ability to read and write in any language	WB HH survey	yes	individual	y	y	n	n
		share of population by highest level of schooling (categorical) (m/f)	DHS	yes	state	y	y	y	y
		share of literate population (m/f)	DHS	yes	state	y	y	y	y
	digital and mechanical know-how	indication if a HH owns a mobile phone	WB HH Survey	yes	HH	y	y	n	n
		share of population who use a mobile phone for financial transactions	DHS	yes	state	y	y	y	y
		share of population who owns a mobile phone	DHS	yes	state	y	y	y	y
		share of households who own a mobile phone	DHS	yes	state	y	y	y	y
		share of households who own a computer	DHS	yes	state	y	y	y	y
		internet usage by frequency (m/f)	DHS	yes	state	y	y	y	y
	social networks (religion)	religion	WB HH Survey	yes	individual	y	y	n	n
	marital status	share of men married or living in union	DHS	no, no data for SSA	state	n	n	n	n
		share of women married or living in union	DHS	no, no data for SSA	state	n	n	n	n
		marital status (categorical)	WB HH survey	no, no data for SSA	individual	n	n	n	n
	age	age of HH head	WB HH Survey	no, no data for SSA	individual	n	n	n	n
		age of manager	WB HH Survey	no, no data for SSA	individual	n	n	n	n
		share of population per age groups in 10-year steps	DHS	yes	state	y	y	y	y
	sex	any female owners	WB EP Survey	no, too many missing values	EP	n	n	n	n
		sex of manger	WB HH Survey	yes	EP	n	y	n	n

		sex of HH head	WB HH Survey	yes	HH	y	y	n	n
		share of female HH heads	DHS	yes	state	y	y	y	y
	awareness and acceptance of productive uses of electricity	-	-	-	-	-	-	-	-
	level of entrepreneurial experience	years of experience of the top manager	WB EP Survey	no, no data for SSA	EP	n	n	n	n
Enterprise Characteristics									
	wealth (income, assets, capital)	total sales of the last fiscal year of the enterprise (in LCU)	WB EP Surveys	yes	EP	n	n	-	y
		total sales in last financial year (Naira)	WB HH Surveys	yes	EP	-	y	n	n
		value of different assets	WB EP Surveys	no, no data for SSA	EP	n	n	n	n
	type/industry	2-digit ISIC 4	WB HH Survey	yes	EP	y	y	n	n
		4-digit ISIC 3.1	WB EP Survey	yes	EP	n	n	y	y
	size	number of full-time employees	WB EP Survey	no, no data for SSA	EP	n	n	n	n
	registration and formality	registration status	WB EP Survey	no, no data for SSA	EP	n	n	n	n
		registration status	WB HH Survey	no, no data for SSA	EP	n	n	n	n
	location of enterprise	location of business at home of owner	WB EP Survey	no, no data for SSA	EP	n	n	n	n
Market Access (general)									
Market Access: Demand									
	population density	Settlement type	Florczyk et al. (2019)	yes	grid-cell	y	y	n	n
		population density	WB HH Survey	yes	grid-cell	y	y	n	n
	urban/rural migration	-	-	-	-	-	-	-	-
	local wealth and purchasing power	share of population by wealth quintile	DHS	yes	state	y	y	y	y
		wealth index gini coefficient	DHS	yes	state	y	y	y	y
	local population/market size	size of locality (categorical)	WB EP Surveys	no, too many missing values	regional	n	n	n	n
	number of people who transit	-	-	-	-	-	-	-	-
	level of competition	-	-	-	-	-	-	-	-
Market Access: Supply/Input factors									
	access to finance (presence of banks, interest rate, required collateral,	share of population with bank account (m/f)	DHS	yes	state	y	y	y	y

	required owners' equity contribution)	Firms using banks to finance investment (% of firms)	WB Development Indicators	yes	national	n	n	y	y
		Firms using banks to finance working capital (% of firms)	WB Development Indicators	yes	national	n	n	y	y
		Lending interest rate (%)	WB Development Indicators	yes	national	n	n	y	y
		community access to bank	WB HH Survey	no, no data for SSA	community	n	n	n	n
		community access to microfinance institution	WB HH Survey	no, no data for SSA	community	n	n	n	n
		experience of HH with borrowing money	WB HH Survey	no, no data for SSA	HH	n	n	n	n
	cost of input factors (other than electricity)	-	-	-	-	-	-	-	-
	presence electricity access	share of HH with electricity	DHS	yes	state	n	n	y	y
		share of population with electricity	DHS	yes	state	n	n	n	n
		national score of the Doing Business Indicator for getting an electrical connection	WB Doing Business	yes	national	n	n	y	y
		access to generator	WB HH Surveys	no, no data for SSA	HH	n	n	n	n
		possession of generator in last FY	WB EP Surveys	no, no data for SSA	EP	n	n	n	n
		availability of electricity in HH	WB HH Survey	yes	HH	y	-	n	n
		nighttime lights	Elvidge et al. (2017)	yes	grid-cell	y	n	n	n
		Access to electricity, urban (% of urban population)	WB Development Indicators	yes	national	n	n	y	y
		Access to electricity, rural (% of rural population)	WB Development Indicators	yes	national	n	n	y	y
		Access to electricity (% of population)	WB Development Indicators	yes	national	n	n	y	y
		available sources of electricity and main source	WB HH Surveys	no, no data for SSA	HH	n	n	n	n
	electricity access capacity (voltage, AC/DC, etc.) quality	Power outages in firms in a typical month (number)	WB Development Indicators	yes	national	n	n	y	y

	(number and duration of blackouts in given time; presence of risk factors for supply)	Value lost due to electrical outages (% of sales for affected firms)	WB Development Indicators	yes	national	n	n	y	y
		Firms experiencing electrical outages (% of firms)	WB Development Indicators	yes	national	n	n	y	y
		Electric power transmission and distribution losses (% of output)	WB Development Indicators	no, too many missing values	national	n	n	n	n
		experience of outages in last FY (dichotomous)	WB EP Surveys	yes	EP	n	n	y	y
		number of outages in an average month in last FY	WB EP Surveys	yes	EP	n	n	y	y
		duration of average outage FY	WB EP Surveys	no, no data for SSA	EP	n	n	n	n
	cost of electrical connection	getting electricity cost	WB Doing Business	yes	national	n	n	y	y
	cost of electricity consumption	national and state electricity tariffs	different sources (see section)	yes	national/ state (only Nigeria)	n	n	y	y
Physical Market Access and Transportation									
	physical access to input goods (e.g. appliances) and services (e.g. maintenance)	indication if a HH owns a television	WB HH Survey	yes	HH	y	y	n	n
		share of households possessing a refrigerator	DHS	yes	regional	y	y	y	y
		share of households possessing a television	DHS	yes	state	y	y	y	y
	level of urbanization	urban cells	Florczyk et al. (2019)	yes	grid-cell	y	y	n	n
		area type (categorical)	European Union GHSL Data Package (2019)	yes	grid-cell	y	y	n	n
		Sector (rural/urban)	WB HH Survey	yes	grid-cell	y	y	n	n
		size of locality (categorical)	WB EP Surveys	no, too many missing values	regional	n	n	n	n
	distance to urban centers	distance to nearest population center (>20.000 inhabitants) (km)	WB HH survey/ WorldCities/ Florczyk et al. (2019)	yes	HH, gird cell	y	y	n	n
		travel time to the next city	Weiss et al. (2018)	yes	grid-cell	y	y	n	n
	presence of roads and distance to roads	distance to nearest major road (km)	WB HH Survey/FERMA/	yes	HH, gird cell	y	y	n	n

			CIESIN et al. (2013)						
		road density	CIESIN et al. (2013)	yes	grid-cell	y	y	n	n
	quality of roads	-	-	-	-	-	-	-	-
	presence of regular market/commercial center	distance to nearest key market centers (km)	WB HH survey/USAID FEWSNET/ Florczyk et al. (2019)	yes	HH, gird cell	y	y	n	n
		community access to market	WB HH survey	no, no data for SSA	community	-	-	-	-
	access to daily public transportation	community access to bus stops	WB HH Survey	no, no data for SSA	community	-	-	-	-
Governance, institutional context and development									
	conductive regulatory environment and policies	-	-	-	-	-	-	-	-
	effort to start a business	national/state Distance to Frontier score of the Doing Business Indicator for the ease of starting a business	WB Doing Business	yes	national /state (only Nigeria)	y	y	y	y
		The rank of ease of starting a business	WB Doing Business	yes	national /state (only Nigeria)	y	y	y	y
		The expected number of days it takes to start a business	WB Doing Business	yes	national /state (only Nigeria)	y	y	y	y
		The expected number of procedures it takes to start a business	WB Doing Business	yes	national /state (only Nigeria)	y	y	y	y
		The expected cost required to start a business (% of income per capita)	WB Doing Business	yes	national /state (only Nigeria)	y	y	y	y
		The expected capital required to start a business	WB Doing Business	yes	national /state (onlyy Nigeria)	y	y	y	y
	level of corruption	-	-	-	-	-	-	-	-
	political relevance and activity of locality (presence of communal administration)	-	-	-	-	-	-	-	-
	level of social security	experience of loss of social services in community (dichotomous)	WB HH Survey	no, no data for SSA	community	-	-	-	-

	availability of subsidies to SMEs	-	-	-	-	-	-	-	-
	country	country	WB EP Survey	yes	-	n	n	y	y
	GDP	GDP PPP	WB Development Indicators	yes	national	n	n	y	y
		GDP PPP/capita	WB Development Indicators	yes	national	n	n	y	y
		GNI PPP	WB Development Indicators	yes	national	n	n	y	y
		GNI PPP/capita	WB Development Indicators	yes	national	n	n	y	y
	gender equality	-	-	-	-	-	-	-	-
	presence of violent conflicts	-	-	-	-	-	-	-	-
	access to water	access to water and quality of supply	WB EP survey	no, no data for SSA	EP	n	n	n	n
	presence of schools	presence of schools (nursery, primary, secondary)	WB HH Survey	no, no data for SSA	community	n	n	n	n
	presence of health facilities	presence of health care facility (health center, hospitals, clinics, doctors, midwifes, dentist, pharmacy)	WB HH Survey	no, no data for SSA	community	n	n	n	n
	investment climate	-	-	-	-	-	-	-	-
	access to cell phone communication	-	-	-	-	-	-	-	-
Other									
	season	-	-	-	-	-	-	-	-
	weather (droughts)	precipitation of the wettest month	WB HH Survey, UC Berkeley (2019)	yes	HH, gird cell	y	y	n	n
		annual mean temperature	WB HH Survey, UC Berkeley (2019)	yes	HH, gird cell	y	y	n	n
		annual precipitation	WB HH Survey, UC Berkeley (2019)	yes	HH, gird cell	y	y	n	n
		occurrence of droughts	Beguería and Serrano (2016)	yes	grid-cell	y	y	n	n
		occurrence of floods	Beguería and Serrano (2016)	yes	grid-cell	y	y	n	n
		mean temperature of the wettest month	WB HH Survey, UC Berkeley (2019)	yes	HH, gird cell	y	y	n	n
Development Program Characteristics									
	training and information provided and quality (BDS), knowledge sharing	-	-	-	-	-	-	-	-

	monitoring progress and satisfaction	-	-	-	-	-	-	-	-
Further potential predictors									
	availability of digital infrastructure and digital know-how	Fixed broadband subscriptions (per 100 people)	WB Development Indicators	yes	national	n	n	y	y
		Mobile cellular subscriptions (per 100 people)	WB Development Indicators	yes	national	n	n	y	y
	(alternative) employment opportunities	share of population who worked in the last 12 months	DHS	yes	state	y	y	y	y
		Vulnerable employment, total (% of total employment) (modeled ILO estimate)	WB Development Indicators	yes	national	n	n	y	y
	Prevalence of poverty	Poverty headcount ratio at $1.90 a day ((2011) PPP) (% of population)	WB Development Indicators	yes	national	n	n	y	y
		Population living in slums (% of urban population)	WB Development Indicators	yes	national	n	n	y	y
		Multidimensional poverty headcount ratio (% of total population)	WB Development Indicators	no, too many missing values	national	n	n	n	n
	Economic activity	New businesses registered (number)	WB Development Indicators	no, too many missing values	national	n	n	n	n
		New business density (new registrations per 1,000 people ages 15-64)	WB Development Indicators	no, too many missing values	national	n	n	n	n
	Energy intensity of an economy (establishedness and efficiency)	GDP per unit of energy use (constant (2017) PPP $ per kg of oil equivalent)	WB Development Indicators	no, too many missing values	national	n	n	n	n
		Investment in energy with private participation (current US$)	WB Development Indicators	no, too many missing values	national	n	n	n	n
		Electric power consumption (kWh per capita)	WB Development Indicators	no, too many missing values	national	n	n	n	n
	Other development indicators	Human Development Index	Global Data Lab	yes	state	y	y	n	n
		Net ODA received (% of GNI)	WB Development Indicators	yes	national	y	y	n	n

Table 19. Variables which were transformed before modelling.

Category	Predictor	Variables found	Source	Model variables	Type of transformation
Household Characteristics					
	Income/level of wealth/assets	share of population by wealth quintile	DHS	share of population in the lowest WQ	-
				share of population in the second WQ	-
				share of population in the middle WQ	-
				share of population in the fourth WQ	-
				share of population in the highest WQ	-
				largest WQ (categorical)	dummy
				average WQ (scale)	weighted average
	number of hh members	share of population per different household size	DHS	avg. number of HH members	weighted average
	shock experience	dichotomous variables on the occurrence of different shocks on community and hh level	WB HH survey	occurrence of floods in the past 3 years (dummy)	disaggregation from community to HH level
				occurrence of droughts in the past 3 years (dummy)	disaggregation from community to HH level
	agricultural activity (precipitation)	share of population occupied in agriculture (m/f)	DHS	share of women occupied in agriculture (m/f)	-
				share of men occupied in agriculture (m/f)	-
				share of avg. population occupied in agriculture	mean
Entrepreneur/HH Head Characteristics					
	level of education (years of schooling, literacy)	share of population by highest level of schooling (categorical) (m/f)	DHS	share of women with secondary or higher education	-
				share of men with secondary or higher education	-
				share of avg. population with secondary or higher education	mean
				share of women with no education	-

Category	Predictor	Variables found	Source	Model variables	Type of transformation
				share of men with no education	-
				share of avg. population with no education	mean
				share women who attended any school	-
				share of men who attended any school	-
				share population who attended any school	mean
		share of literate population (m/f)	DHS	share of literate women	-
				share of literate men	-
				share of literate avg. population	mean
	digital and mechanical know-how	share of population who use a mobile phone for financial transactions	DHS	share of women who use a mobile phone for financial transactions	-
				share of men who use a mobile phone for financial transactions	-
				share of overall population who use a mobile phone for financial transactions	mean
		share of population who owns a mobile phone	DHS	share of women who own a mobile phone	-
				share of men who own a mobile phone	-
				share of overall population who owns mobile phone	mean
		internet usage by frequency (m/f)	DHS	women who ever used the internet	-
				men who ever used the internet	-
				avg. population who ever used the internet	mean
	age	share of population per age groups in 10-year steps	DHS	average age of population	weighted average
				share of population aged 30-49	sum
	social networks (religion)	religion	WB HH Survey	religion dummies	dummy
Enterprise Characteristics					

Category	Predictor	Variables found	Source	Model variables	Type of transformation
	wealth (income, assets, capital)	total sales of the last fiscal year of the enterprise (in LCU)	WB EP Surveys	natural logarithm of total sales of the last fiscal year of the enterprise (USD 2020)	natural logarithm, transform currency, adjust for inflation
		total sales in last financial year (Naira)	WB HH Surveys	total sales of the last financial year in USD 2020	transform currency and adjust inflation
	type/industry	4-digit ISIC 3.1	WB EP Survey	3-digit ISIC 3.1	aggregate
				2-digit ISIC 3.1	aggregate
				categorical ISIC 3.1	aggregate
Market Access: Demand					
	population density	area type (categorical)	European Union GHSL Data Package (2019)	aera type dummies	dummy
	local wealth and purchasing power	share of population by wealth quintile	DHS	share of population in the lowest WQ	-
				share of population in the second WQ	-
				share of population in the middle WQ	-
				share of population in the fourth WQ	-
				share of population in the highest WQ	-
				largest WQ (categorical)	dummy
				average WQ (scale)	weighted average
Market Access: Supply/Input factors					
	access to finance (presence of banks, interest rate, required collateral, required owners' equity contribution)	share of population with bank account (m/f)	DHS	share of women with bank account	-
				share of men with bank account	-
				share of overall population with bank account	mean
	electricity access capacity (voltage, AC/DC, etc.) quality (number and duration of blackouts in given time; presence of risk factors for supply)	Power outages in firms in a typical month (number)	WB Development Indicators	power outages in a typical month	missing values were substituted by country average obtained by the WB Development Indicators
		number of outages in an average month in last FY	WB EP Surveys		
Physical Market Access and Transportation					
	level of urbanization	area type (categorical)	European Union GHSL Data Package (2019)	aera type dummies	dummy
Governance, institutional context and development					

Category	Predictor	Variables found	Source	Model variables	Type of transformation
	country	country	WB EP Survey	SSA region dummy	dummy
	employment	share of population who worked in the last 12 months and are currently	DHS	share of women who worked in the last 12 months	-
				share of men who worked in the last 12 months	-
				share of avg. population who worked in the last 12 months	mean

13.5.1 Appendix: Transforming Wealth Quintiles

It was considered to transform the dominant WQ variables further to summarize the WQ to two groups since rather than indicating the WQ a HH belongs to, the variable just indicates to most likely WQ a HH belongs to. This probability (i.e. the share of the predominant WQ in the region of a HH) was especially weak for the middle quintiles (see Table 20 for the data form the HH survey). Especially for the EP survey the EP might not necessarily say something about the actual wealth of the HH because areas are quite heterogenous, i.e. have relatively equal shares of all WQ, which can be seen when looking at the average probability of an EP to belong to the WQ that is dominant in its region (see Table 21). Summarizing, e.g. WQ 1+2 and W3+4+5 would have increased the probability of a HH to belong to the assigned wealth quintiles (see Table 20). Since these variables did not turn out to be more useful that the other ones, they were not applied in the final models.

Table 20. Increased classification accuracy of HH from summarizing dominant WQ in the HH sample.

	Avg. probability of a HH to belong to its assigned dominant WQ
WQ5	48%
WQ4	33%
WQ3	31%
WQ2	34%
WQ1	49%
	Avg. probability of a HH to belong to its assigned dominant WQ
WQ 5+4+3	92%
WQ 1+2	72%

Table 21. Increased classification accuracy from summarizing dominant WQ in the EP sample.

	Avg. probability of an EP to belong to its assigned dominant WQ
WQ5	54%
WQ4	29%
WQ3	30%
WQ2	31%
WQ1	39%
	Avg. probability of an EP to belong to its assigned dominant WQ
WQ 5+4+3	92%
WQ 1+2	62%

13.6 Appendix: Missing values

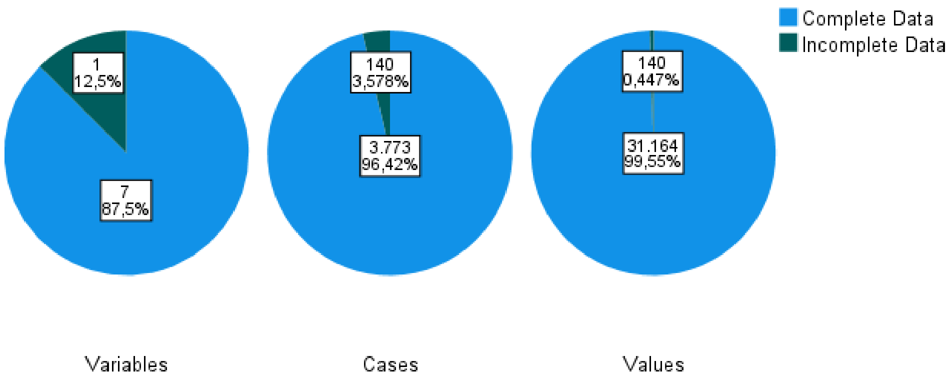

Figure 34. Summary of missing values for the connection regression.

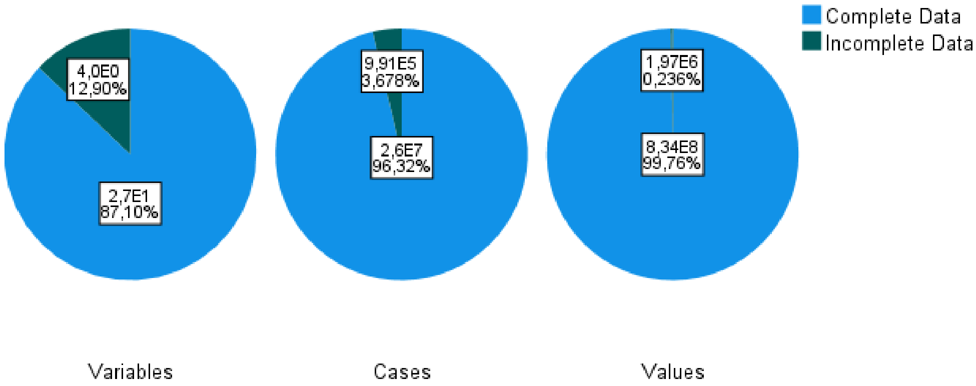

Figure 35. Summary of missing values for the uptake regression.

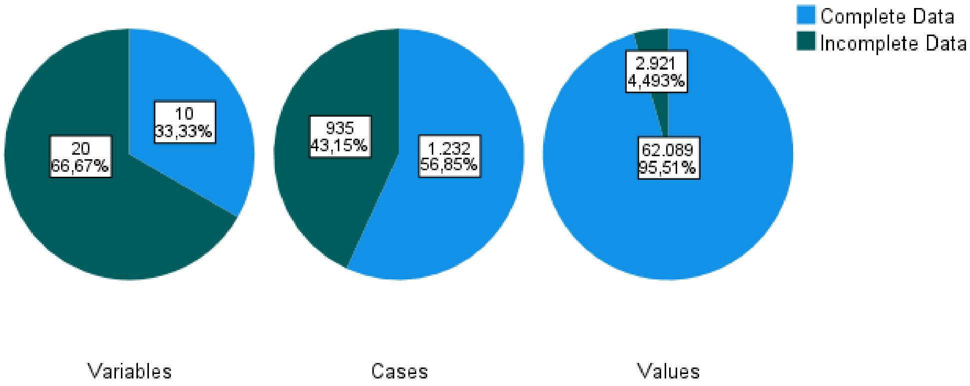

Figure 36. Summary of missing values for the EP survey.

13.6.1.1 Imputing missing values

The imputation of values was attempted using the multiple imputation algorithm in SPSS as it is generally accepted as the most accurate form of imputing values (Royston & Division, 2004). This method is a regression-based approach to estimate missing values based on preselected predictors. However, since there was little certainty about predictors yet, all potential independent variables had to be assumed to be relevant and used as predictors, which the algorithm could not handle. In general, it is possible to set restrictions to the value that a variable can have (e.g. so that no negative values for percentages are imputed etc.). However, even after increasing the maximum iterations of the algorithm form the base value of 100 to 2500 draws, the algorithm could not converge and restrictions on some variables had to be lifted. The resulting imputed values were sound for most variables but some of them had illogical values. It was attempted to run regressions anyway to compare results, using the same variables as with the case-wise exclusion of missing values and it was found to lead to similar results with coefficients varying by around 10% and minor improvements of the R^2. However, the assumptions of the underling regression analysis were not fully accessible, but it was most likely very similar to the model regressions anyway, adding not much more insight but adding some illogical imputed values to the data. It was therefore decided that the imputation was not necessary but would lead to an unnecessary loss of transparency and was therefore discarded. At later stages of narrowing down the number of variables for the final model, it was checked again if imputing values was feasible and if it would benefit the model. The algorithm worked better and some of the initially intended restrictions on variable values could be lifted but the imputed values were still partly illogical or simply not in line with the original values.

13.7 Appendix: Representativeness and ranges of Variables

Table 22. Range of values of variables in the final uptake model.

	Mean	Median	Minimum	Maximum
Share of women employed in state	66.77	71.10	37.10	94.00
Share of men employed in state	85.06	85.90	66.00	98.20
Number of HH members	6.17	6.00	1.00	33.00
Share of population aged 30-49 in state	20.83	21.40	14.90	30.00
HDI	0.55	0.60	0.34	0.67
Share of population with access to electricity	59.28	55.40	20.70	98.90
Population density	2016.35	484.00	12.00	56965.00
Number of urban cells	22.04	7.51	0.00	91.25
Travel time to the next city	23.27	15.30	0.00	476.22
Distance to nearest major road	5.15	2.20	0.00	59.30
Distance to nearest population center	22.40	17.70	0.20	155.40
Distance to nearest market	64.68	58.40	0.40	227.00
Starting a Business days	33	28	6	91
settlement type=10.0	-	-	1	2
settlement type=11.0	-	-	1	2
settlement type=12.0	-	-	1	2
settlement type=13.0	-	-	1	2
settlement type=21.0	-	-	1	2
settlement type=30.0	-	-	1	2
Presence of EP in HH	-	-	1	2
Access to electricity in HH	-	-	1	2
Agricultural activity in HH	-	-	1	2
Presence partner of HHH	-	-	1	2
Sex of HHH	-	-	1	2
School attendance of HHH	-	-	1	2
Sector	-	-	1	2
dom_WQ_new=1.0	-	-	1	2
dom_WQ_new=2.0	-	-	1	2
dom_WQ_new=3.0	-	-	1	2
dom_WQ_new=4.0	-	-	1	2
dom_WQ_new=5.0	-	-	1	2
HH ownership of mobile phone	-	-	1	2
HH ownership of TVs	-	-	1	2
Experience of drought in community (3y)	-	-	1	2
Experience of flood in community (3y)	-	-	1	2

Table 23. Range of values of variables in the final connection model.

	Mean	Median	Minimum	Maximum
Sales in last FY in USD	2426.18	997.23	0.00	99722.99
share of HH in hi WQ in state	22.83	18.50	1.50	75.00
Number of urban cells	25.45	11.18	0.00	91.25
Literacy of manager	-	-	1	2
Agricultural activity in HH	-	-	1	2
HH ownership of mobile phone	-	-	1	2
Experience of drought in community (3y)	-	-	1	2
Access to electricity in HH	-	-	1	2

Table 24. Range of values of variables in the final sales and electricity consumption model.

	Mean	Median	Minimum	Maximum
ln Sales in last FY (USD)	9.33	9.47	4.82	11.78
ln El. consumption last FY (kWh)	6.52	6.72	-1.98	14.45
Share of female HHH in state	27.14	28.10	1.20	49.40
Share of women with secondary education in state	24.45	13.50	0.30	89.20
GNI, PPP (constant 2017 international $)	76,921,103,770.07	37,621,726,854.24	2,934,091,478.03	936,915,193,284.25
National net ODA received (% of GNI)	9.70	8.65	0.45	18.43
Power outages in firms in a typical month (number)	10.43	4.50	0.60	32.80
Starting a Business Dtf	73.54	74.40	49.20	90.00
Getting Electricity score	44.70	43.90	29.40	78.10
Share of pop. without education	15.69	4.20	0.10	76.95
National lending interest rate (%)	13.30	15.68	4.98	29.75
share of pop. in lowest WQ in state	12.39	4.30	0.00	55.30
share of pop. in low-mid WQ in state	14.21	13.10	0.00	39.50
share of pop. in middle WQ in state	15.07	17.30	0.00	35.30
share of pop. in mid-hi WQ in state	21.61	21.50	3.40	39.60
share of pop. in highest WQ in state	36.71	32.90	1.50	95.50
size of locality=capital city	-	-	0	1
size of locality=over 1 mio	-	-	0	1
size of locality=250.000 to 1 mio	-	-	0	1
size of locality=50.000to250.000	-	-	0	1
size of locality=less than 50.000	-	-	0	1
SSA region=Central Africa	-	-	0	1
SSA region=West Africa	-	-	0	1
SSA region=Southern Africa	-	-	0	1
SSA region=East Africa	-	-	0	1
dominant wealth quintile=1.0	-	-	0	1
dominant wealth quintile=2.0	-	-	0	1
dominant wealth quintile=3.0	-	-	0	1
dominant wealth quintile=4.0	-	-	0	1
dominant wealth quintile=5.0	-	-	0	1

Table 25. Distribution of EP in the HH Sample and the EP Sample according to their 2-digit ISIC Rev 3.1 code.

ISIC 3.1 (2-digit code)	HH	EP
Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods	42%	39%
Other personal service activities	14%	-
Hotels and restaurants/Food and beverage services	7%	6%
Land transport and transport via pipelines	7%	1%
Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel	6%	8%
Manufacture of wearing apparel; dressing and dyeing of fur	5%	10%
Manufacturing of food	3%	4%
construction	2%	2%
Manufacture of furniture; manufacturing n.e.c.	1%	6%
Manufacturing of wood	1%	2%
Manufacturing of fabricated metal products	1%	4%
Wholesale trade and commission trade, except of motor vehicles and motorcycles	1%	5%
Publishing, printing and reproduction of recorded media	-	3%
Post and telecommunications	-	1%
Computer and related activities	-	2%
Other	10%	8%

13.7.1 Comparison of dependent variables with literature

Nagler & Naudé (2017) find different shares of EP-HH in different SSA countries, varying from 17 to 62 percent. In urban areas, the shares are more stable, varying from 35 to 68%. In tendency, they find the share of EP to be higher in urban than in rural areas. Other than that, there is little information on the share of HH with EP in literature.

Concerning electrification rates, there is only information from case studies found in literature. These indicate a connection rate of 37%-90% about 1-3 years after electrification but the time was not always given. More precisely, the values found were 37, 38, 40, 60, 61.8 and 62% respectively (Bensch et al., 2016; Kariuki et al., 2016; Lenz et al., 2017; Mayer-Tasch et al., 2013). Some studies indicated the reasons of those enterprises which did not connect. Usually these reasons were affordability of either the connection fee, the consumption cost or the cost of appliances and only few EP sated they did not need electricity. If all enterprises which want to connect could connect, the electrification rates would be much higher, more precisely, instead of 38%, 75% of EP would connect, the rate from 60% would increase to 100% and the rate of 62% would be at 90% instead. Some studies provide logit or probit models to find determinants for connection, but they use some enterprise-specific variables that are not available for the projection to all of SSA. Similar information is available for households which provides similar shares (Bernard, 2012; Golumbeanu & Barnes, 2013; Lenz et al., 2017; Williams & Jaramillo, 2018).

Table 26 compares the levels of electricity consumption found in the EP Survey with findings of similar business types from other literature.

Table 26. Comparison of electricity consumption in kWh per business type according to ISIC and literature.

Annual electricity consumption (kWh)										
EP Surveys			Mayer-Tasch et al. (2013)		Kirubi (2018)		Banerjee (2017)		Bensch et al. (2016)	
Business type	**Mean value**	**Median value**	**Mean value**	**Business type**	**Mean value**	**Business type**	**Mean value**	**Business type**	**Mean value**	**Business type**
Manufacturing	5659	1720	840	Non-reliant manufacturing	3300	Carpentry and workshops	15120	Mechanical workshop (welding, grinding, drilling)	3144	Manufac-turing
			3624	Reliant manufacturing			3600	Bakery		
Construction	5402	2075	-	-	-	-	-	-	-	-
Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods	5085	1526	240	Commerce	473	Retail and repair shops	1620	Small shops	1296	Trade
Hotels and restaurants	7578	3785	2628	Bars	180	Small tea/food café	-	-	-	-
					2880	Bars, lodging, hotel	-	-	-	-
Transport, storage and communications	6203	1378	-	-	-	-	-	-	-	-
Real estate, renting and business activities	6483	2092	-	-	-	-	5400	Media center	-	-
-	-	-	720	Reliant service	8800	Grain mills	3600	Milling or grinding	1500	Service
-	-	-	360	Hairdressers	-	-	-	-	-	-
-	-	-	-	-	2271	Petrol station and welding garage	1800	Petrol station	-	-

13.8 Appendix: Descriptive Statistics of Uptake Predictors and general EP characteristics

As uptake is mostly a question of choice made on the HH basis, this section will mainly focus on the findings in the HH sample to uncover differences between households and household heads with EP (EP HH) and households without EP (non-EP HH).

13.8.1 Household characteristics

Across the 4979 HH in the sample (weighted 26,957,052 HH), 60% indicated that at least one on the HH members employed at least one EP. However, since some of the HH operate several EP, the number of EP in the sample is 3913 (or weighted 23,096,939). More specifically, of the HH with enterprise (EP HH), 68% have one EP in the household, 25% have two, another 5% have three and the remaking 2% have more than three (up to eight EP). As already expected in the literature review, a larger HH tends to favor EP activity. While the average HH size is 6.17, it is 6.5 for HH with EP and 5.70 for HH without EP.

74% of the EP are only operated by one person, 98% employ less than five people. Few EP earned more than 100,000 USD/year so that filtering for these two criteria still left 22,623,033 weighted and 3,820 unweighted cases. 84% of EP were operational during the entire year. Furthermore, 70% of urban HH have an EP and only 54% or rural HH have an EP. The distribution of EP HH and non-EP HH across wealth quintiles seems to be roughly the same except for a slightly higher share of EP-HH in the lowest quintile (Figure 37).

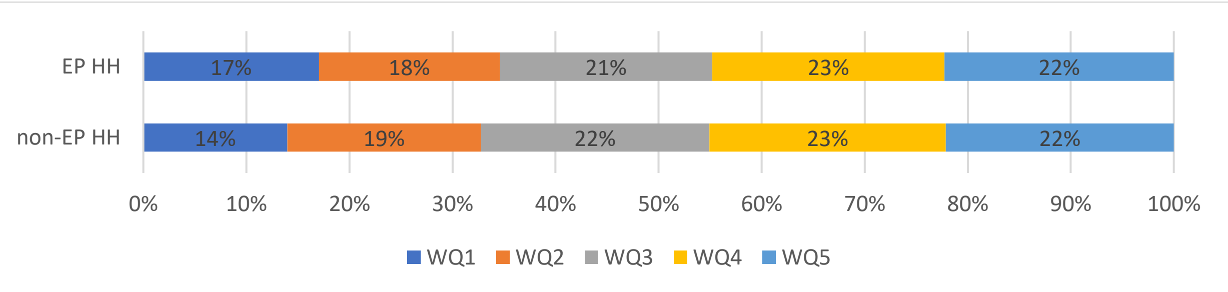

Figure 37. Comparison of wealth distribution between EP HH and non-EP HH.

When looking at the wealth distribution between rural and urban areas in Figure 38, it can be seen that the prevailing pattern is a tendency of more EP HH in the poorest WQ than non-EP HH and since rural areas tend to be less wealthy than urban, the share of EP HH in the lowest WQ is larger in rural areas than the share of EP urban areas, which are mostly located in the highest WQ. However, in urban areas, WQ2 and WQ3 seem to have more EP activity while in rural areas, it is only the lowest WQ in which EP HH are more present than non-EP HH. Perhaps for rural HH, subsistence agriculture is still a relatively more important strategy to cope with poverty as compared to urban areas.

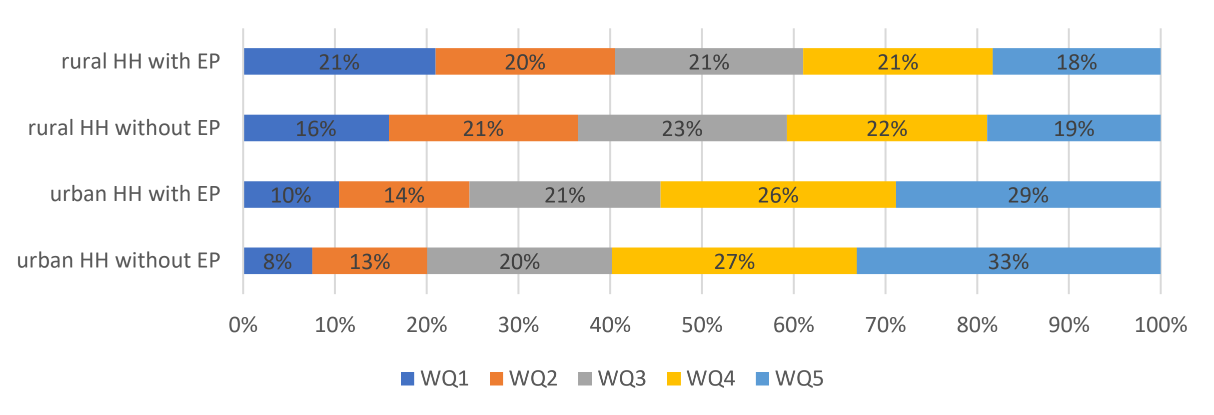

Figure 38. Comparison of wealth distribution of EP HH and non-EP HH in rural and urban areas.

Since the income of a HH is directly related to the income from the EP, it is difficult to make statements of the level of income on the propensity of EP uptake but it can be said that the average total income as calculated in 6.3.2 per HH member is 60% higher for EP-HH than for non-EP HH and no-EP income for both EP.

Looking at different conditions as given in Table 27 brings some perspective to this finding. Total income per HH member is about the same for EP HH and non-EP HH when both have access to electricity (1,395 and 1,249 USD respectively) but for not connected HH, EP-HH tend to earn twice as much (500 USD) than non-EP HH (243 USD). The average income to HH where the HHH did not attend school is about the same for non-EP as for EP HH (520 and 457 USD) while a higher education leads to a twice as high income to EP HH as for non-EP HH (1160 and 723 USD). While income is about the same for EP and non-EP HH in rural areas (683 and 729 USD), EP-HH earn much more in urban areas than HH without EP (1655 and 187 USD), despite the wealth distribution. There are further interaction effects such as a higher share of school attendance and higher connection rates in urban than in rural areas (87% vs. 75% for school attendance and 84% vs. 43% for electricity access).

Finally, some things can be said about the effect of agriculture on EP income. Of the entire sample, 70% of HH indicated that at least one member of the HH cultivated any crops in the last year while it was only 34% HH in urban areas and 87% in rural areas. Of those HH which cultivated crops, 56% also had an EP while those HH without agricultural activities, 67% had their own enterprise. The findings about the relationship between agricultural activity and EP was ambiguous in the literature review but in this sample, it is rather negatively associated with EP activity. Concerning the association of HH income and agricultural activity, income of HH who have income from agriculture but not from EP have the lowest income (201 USD per year and member), followed by HH who have both (591 USD). Next come EP HH which do not rely on agricultural incomes (1.953 USD) but the highest earing HH are by far those who indicate to have neither income from an EP nor form agricultural activity with 4,491 USD/member/year.

Table 27. Patterns in average total income per HH member.

	average total annual HH income per HH member in USD 2020	
	HH without EP	**HH with EP**
Urban	187	1655
Rural	729	683
Access to electricity	1249	1395
No access to electricity	243	499
School attendance of HHH	723	1160
No school attendance of HHH	520	457
Agricultural activity	201	591
No agricultural activity	4491	1953

From this descriptive analysis, there is already some evidence on the impact of lower wealth and higher urbanization to favor the uptake of EP, indicating the prevalence of push-businesses. However, the connection is not straightforward as covariates were found to interact in quite complex ways.

13.8.2 Household head and entrepreneur characteristics

To make HH comparable, the focus here is only on the characteristics of the household head (HHH) or the manager of a HH-EP. In 48% of EP, it is the HHH him-/herself who is managing the EP, in 37% it is the spouse and in further 11% the own child.

Literacy and school attendance rates of the household head are higher amongst the EP HH as compared to non-EP HHH with 84% having attended school in EP-HH and only 71% in non-EP HH. Literacy is at 80% for EP HHH and 67% of non-EP HHH. There is a noteworthy relationship between sex and education with rates being about 20-30 percentage points higher for male HHH as well as male managers as compared to female HHH and female managers. While 86.8% of male managers in the HH survey enterprises are literate, only 68.1% of female owners are (Table 28).

Table 28. Literacy of EP managers and HHH by sex.

	Literacy of manager	**School attendance of manager**	**Literacy of HHH**	**School attendance of HHH**
male	87%	90%	80%	82%
female	68%	76%	51%	62%

But there is even more to uncover about the relationship of sex and entrepreneurship. 81% of HHH in the sample are male. Of those, 61% have an EP while of the female headed HH only 53% have an EP. Of all EP with a male HHH, only half are also operated by a man while of those EP of HH with a female HHH, 85% are also operated by a woman. Female household headship thus seems to be quite indicative for female entrepreneurship (see Table 29).

Table 29. Relationship between sex of HHH and sex of EP manager.

		Sex of HHH	
		male	**female**
Sex of EP manager	male	44%	2%
	female	41%	13%

The result is that 51% of EP are being managed by a woman. This share tends to increase as wealth increases (Table 30). This tendency was unclear in literature.

Table 30. Share of female entrepreneurship in HH Survey.

	Dominating WQ in state					
	1	**2**	**3**	**4**	**5**	**Total**
Share of businesses with female manager	42.6%	48.3%	55.6%	60.3%	53.8%	51.1%

Interestingly, there is a pattern of rather consistent entrepreneurial activity amongst female headed HH of a little more than 50% while entrepreneurial activity amongst male headed HH is at only 41% for those without partner and at 63% for those with partner. This is a typical example for an interaction effect, which can lead to unintuitive distortions of coefficients in the regression models if not treated properly (see Appendix 13.16.6 for more information).

In the EP surveys, the indication if there are any female owners was given for 2205 cases, of which 36% of EP confirmed that there was at least one female owner. The share also tends to increase with increasing wealth with WQ4 being an exemption (see Table 31). However, given the low number of cases for WQ4, this exception could be evaluated as not representative.

Table 31. Share of female entrepreneurship in EP Surveys.

	Dominant WQ in state of residence					
	1	**2**	**3**	**4**	**5**	**Total**
Share of businesses with female owners	25%	26%	33%	16%	43%	36%

Finally, the average age for owners of both sexes is 40 but the minimum for girls is 10 and for boys it is 12 years of age. The oldest male owner is 99 years old and the oldest female owner 96. The minimum age increases with increasing wealth as can be seen in Table 32. The youngest entrepreneur amongst those in areas where most people belong to WQ5 being 17 years old. This is an indication for a somewhat appropriate classification of wealth quintiles as child labor is especially associated with poverty. For the maximum age, however, there seems to be no such trend. The average age only slightly increases with increasing wealth.

Table 32. Age distribution across wealth quintiles.

		Age of the youngest entrepreneur	Age of the oldest entrepreneur	Average age
Dominating WQ in state	1	10	80	40
	2	12	85	37
	3	10	96	40
	4	16	99	43
	5	17	85	43

Overall, findings offer some evidence that confirm the conclusion by the literature, which finds women to be more engaged in the poorer push-businesses. This impression is enhanced when looking at the types of businesses women tend to engage in as opposed to men, as explained in the analysis of business types by ISIC code in Appendix 0.

13.8.3 Market Access

According to literature, a major predictor of access to markets for both demand and supply are the level of urbanization and population density. 69% of HH in the sample are situated in rural areas. Indeed, while the average share of HH with EP in the sample is 60%, it is much higher in urban areas with 71% and lower in rural areas with 54%. The population density is also noticeably higher for EP-HH (2300 ppl/km^2 as compared to 1599 ppl/km^2 for non-EP HH). This is in line with the finding that most of the EP are located in urban areas.

Concerning access to finance, the share of EP with HH is about ten percentage points higher in communities which have access to either a bank or a microfinance institute (Table 33).

Table 33. Access to finance in terms of banks and microfinance institutions and EP Uptake in HH Survey.

	Share of HH with EP
Communities with access to banks	68%
Communities without access to banks	57%
Communities with access to microfinance institution	68%
Communities without access to microfinance institution	58%

Another, more specific predictor concerning market access, is the access to input factors, for which access to electricity is of special interest in this research. As already pointed out in section 6.10.1, 56% of households indicate having some sort of access to electricity (83% of urban and 37% of rural HH). While it was found that amongst urban HH, the presence of electricity did not change the rate up EP uptake much (71% as compared to 70% on average), electrified HH in rural areas show to have a noticeably higher uptake rate of 61% as compared to the average 54%. Even though access to electricity has some predictive power for EP uptake, this correlation alone does not yet allow of a clear interpretation of causality. Perhaps the presence of electricity in an area allowed for more EP uptake because it allows for productive uses, perhaps existing EP decided to get connected to improve their business or maybe the electricity is not even used for the EP but both are merely associated with other factors such as wealth or education, since the connection refers to the household, not the EP itself.

Finally, it was found that different predictors have been found to be associated to different degrees and sometimes different directions with different types of businesses as identified by 2-digit ISICs. As this information could not be used in the model, the closer investigation on this issue can be found in Appendix 0.

13.9 Appendix: Descriptive Statistics of Connection

This section will investigate the main effects of the available predictors identified in the literature on the connection behavior of EP in a descriptive way. As seen in

Table 34, indications have been found to be quite straightforward for the connection rate and will be briefly explained in this section, while Appendix 13.11 will then have a closer look at predictors associated with the actual level of consumptions. The present section also explains the ambiguous predictors pointed out in section 7, i.e. when it is not known to which extent a predictor predicts the availability of electricity supply or the motivations of an EP to demand electricity under perfect supply conditions to understand in how far this might limit the interpretability.

Sales of EP with connection are on average higher than those of EP without, which is in line with literature findings even though it cannot be said for sure in how far the one is a cause for the other. This correlation also means that it can be expected that there is going to be some overlap of covariates which drive performance that also increase the probability to get an electrical connection, which can dilute the explicability of clear causal effects as explained in section 7.

There is a slight but homogenous trend that the share of EP with electricity increases as the share of population in the higher WQ increases in the state of residence of an EP. Concerning market access, more urbanized areas are clearly associated with a higher rate of connection and the number of urban cells as well as population density are higher for EP with electrical connection. This is in line with the finding that higher WQ, higher income and urbanization tend to be associated and support the idea that connectivity is more associated with opportunity-driven business rather than needs-driven businesses. On the other side, it could also mean that the supply or electricity is simply better in these areas and it is therefore easier for EP to connect. Without qualitative knowledge about this, it will just have to be assumed that these covariates drive demand rather than supply as stated In Section 7.

Another rather intuitive finding in line with this narrative is that literacy and school attendance rates as proxies for education and knowledge are positively associated with electrified EP, as connection rates are about 25% higher for literate managers and those who attended school. Mobile phone ownership, which can be seen as an approximation for digital literacy and wealth, is also positively associated with connection. EP whose HH possesses at least one mobile phone have a 28% higher connection rate than EP without mobile phones in the HH. Concerning sex of the manger, however, virtually no difference in the connection rate could be found. Finally, the occurrence of droughts in the community have also been found to be associated with connectivity of EP. Different connection rates have furthermore been found for different ISICs, on which more details can be found in Appendix 0.

Table 34. Differences between EP with electricity and EP without electricity.

		EP without electricity	EP with access to electricity
Total share of EP		38%	62%
share of pop. in lowest WQ in state		25%	11%
share of pop. in low-mid WQ in state		23%	15%
share of pop. in mid WQ in state		21%	21%
share of pop. in mid-hi WQ in state		18%	25%
share of HH in hi WQ in state		13%	29%
Sales in last FY in USD		1332.40	2897.09
Number of urban cells		9.18	35.39
Population density		745	3944
Sector	1. Urban	14%	86%
	2. Rural	53%	48%
Literacy of manager	Yes	33%	67%
	No	57%	43%
School attendance of manager	Yes	33%	67%
	No	60%	40%
Sex of manager	Male	40%	60%
	Female	37%	63%
HH ownership of mobile phone	Yes	33%	67%
	No	51%	49%
Experience of drought in community (3y)	Yes	66%	34%
	No	32%	68%

13.10 Appendix: Descriptive Statistics of Performance

Figure 39 and Figure 40 show the frequency distributions of sales for the HH and the EP sample respectively, once on a normal scale, once on a logarithmic scale (N=3,913 and N=1,753). On the logarithmic scale, the values are approximately normally distributed, otherwise, they are strongly skewed to the right side. The graphs also illustrate again that the level of sales is much higher in the EP Surveys than in the HH Survey. In the HH sample, the sales have a mean of 2,426 and a median of 997 USD as well as a standard deviation of 5,741. In the EP sample, the sales have a mean of 25,087 and a median of 13,646 USD at a standard deviation of 27,596.

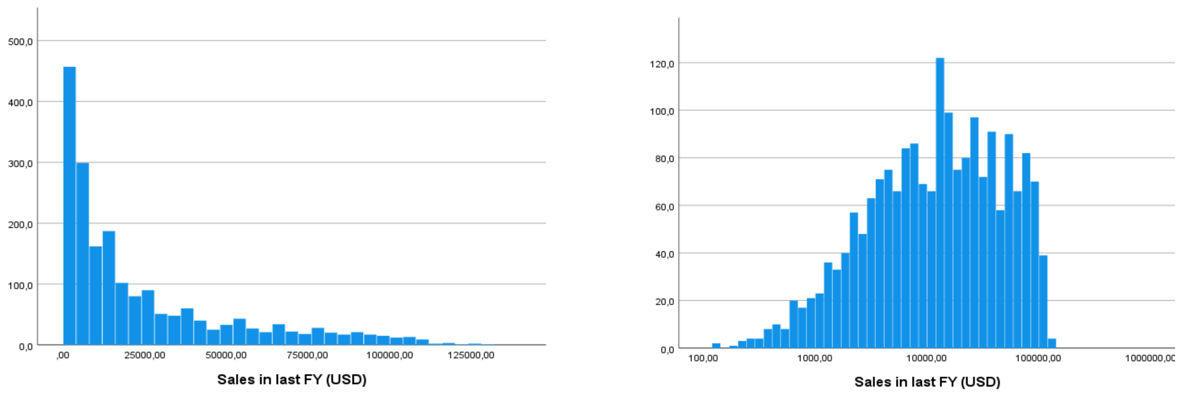

Figure 39. Frequency distribution of sales of HH sample on a normal and a logarithmic scale.

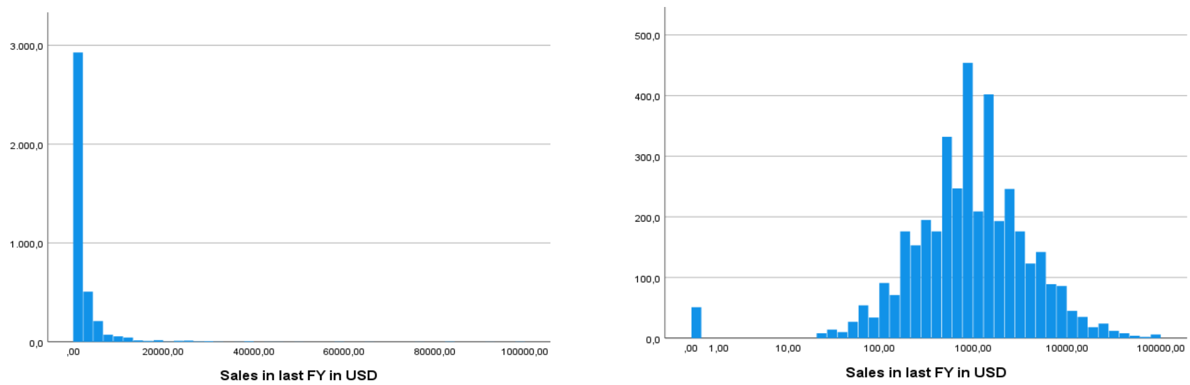

Figure 40.Frequency distribution of sales of EP sample on a normal and a logarithmic scale.

There is some consistency of actual wealth with the wealth quintiles (see Table 36 at the end of the section for a summary of all variables). While uptake seemed to have been the strongest in the lowest WQ, the performance of EP tends to better when located in areas with more people in the upper WQ with a slight dip of the fourth quintile and rather high level of sales in the lowest quintile. For the EP surveys, the sales per year per dominant WQ follow a similar pattern but with more extreme values for the exceptions of WQ1 and WQ4. Looking at the standard deviations and frequency distributions of sales values (on a logarithmic scale) (Figure 41 and Figure 42) makes visible that there is also quite a spread of sales within each of the wealth quintiles. These blurry findings might be due to poor representativeness of the variable as mentioned in section 6.

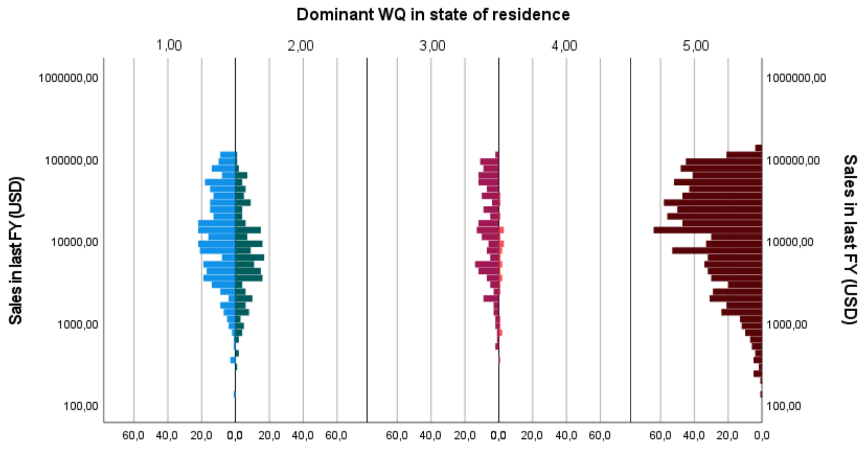

Figure 41. Frequency distribution of sales per dominant WQ category of EP sample.

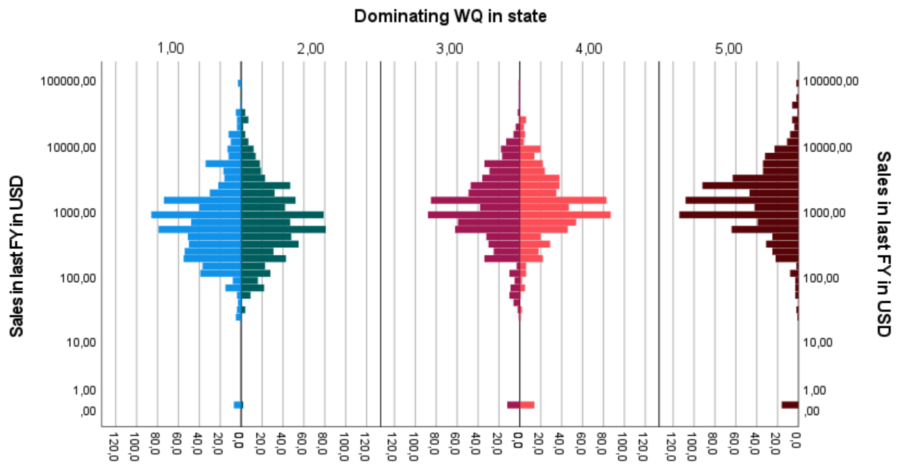

Figure 42. Frequency distribution of sales per dominant WQ category of HH sample.

As can be seen in Table 35, the differences in average sales of EP in each country mostly align with the distribution of WQ per country, visible exceptions only being The Gambia and Namibia, which have low sales despite a high share of the upper WQ. This pattern also carries on when looking at sales of different ISIC (see Appendix 0). Indeed, comparing the standard deviations of sales for each country and of each ISIC indicated that the deviance was much smaller within country groups (avg. standard deviation of all countries of 12,406) than within ISIC groups (average standard deviation of 17,826), indicating a better explication of sales level by country than by ISIC. It is hoped that by factoring in the country SSA region, this can capture some of the explanation.

Table 35. Wealth and sales per country in the EP sample.

		WQ1	WQ2	WQ3	WQ4	WQ5	Sales in last FY (USD)
Central Africa	Cameroon	0%	3%	10%	34%	53%	28,666
	DRC	23%	23%	21%	18%	14%	14,809
West Africa	Nigeria	18%	18%	18%	21%	24%	13,701
	Gambia. The	0%	3%	10%	34%	54%	6,609
	Burkina Faso	3%	6%	7%	9%	76%	30,790
	Ivory Coast	28%	29%	23%	14%	6%	10,687
Southern Africa	Namibia	1%	10%	14%	25%	50%	11,979
	Swaziland	13%	19%	20%	23%	26%	21,733
	Mozambique	17%	19%	17%	20%	26%	13,108
	Zimbabwe	10%	10%	12%	29%	39%	51,892
East Africa	Uganda	6%	4%	3%	7%	80%	27,928
	Rwanda	4%	5%	7%	17%	67%	40,512
	Kenya	9%	15%	16%	22%	38%	31,241

Concerning entrepreneur characteristics, the strongest predictor for sales is the level of education. EPs in the EP Surveys whose owners have no education have only about a third of the sales than the average. The HH Survey also shows that school attendance and literacy of the owner are associated with about 100% higher sales. The HH Survey data further confirms findings form literature that most female operated businesses earn less than male operated businesses. Conversely, there is no such evidence from the EP Survey data. The lesser impact could be explained by the generally higher level of wealth in the EP surveys, which makes it less necessary for women to engage in survival-type businesses, which they were often found to operate in (Lecoque & Wiemann, 2015).

A deeper insight on this issue can be seen in the comparison of ISIC (see Appendix 0). For example, there is a tendency found in both surveys of women to engage in more businesses with a generally lower income such a clothing and food manufacturing, as indicated in literature but they also tend to earn less across different business types as compared to their male counterparts.

Concerning predictors of access to markets, the only indicator given by the EP Survey, the size of locality, has a tendency to be negatively correlated with sales, which is a counterintuitive finding but even consistent over different levels of education, different WQ and ISIC. The average standard deviation of values across the different size categories is rather high compared to the ones for education or WQ. However, the tendency that smaller places have a higher consumption remains. This effect can, however, also occur due to weakness of the sample rather than an actual trend due to the few cases in the smallest category. Even though not implementable in the EP model, it is interesting to look at the different average incomes of EP in rural and urban areas as indicated in the HH survey to check if this finding can be confirmed or even explained. Both EP and non-EP income is much higher in urban areas. This is in line with literature and seems to contradict the findings from the EP surveys. Even the same businesses of the same type tend to earn more in urban than in rural areas. However, since there is no differentiation for different town sizes, there might be effects that are not captured by only looking at the dichotomous urbanization variable, e.g. a larger population does not only mean higher demand but also potentially stronger competition and a higher need to make lower, competitive prices, which can decrease sales.

With respect to access to electricity, section 13.9 already pointed out the positive relationship with sales found in the HH Survey. But also in the EP Survey, a positive correlation was found between sales and electricity consumption (Figure 43) and sales and outages (Figure 44) on a logarithmic scale. As will be discussed in the next section, the causality of these correlations is not fully clear. However, it is already useful to know that these variables can serve as predictors for each other.

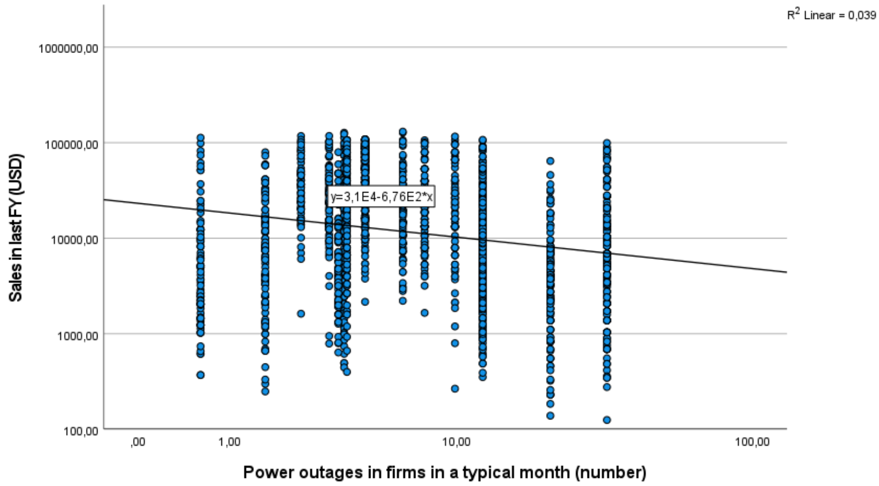

Figure 43. Relationship between sales and outages in the EP sample.

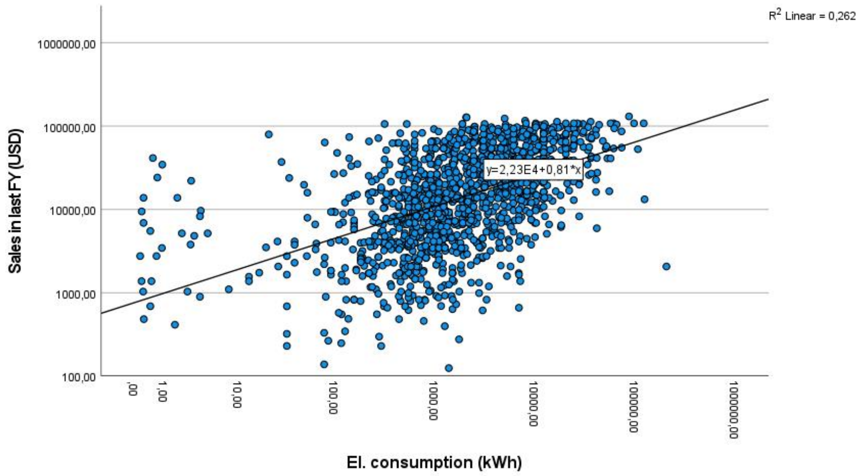

Figure 44. Relationship between sales and electricity consumption in the EP Survey.

Table 36. Patterns in sales levels in HH and EP Surveys.

		Sales in last FY (USD) HH Survey	Standard Deviation HH Survey	Sales in last FY (USD) EP Survey	Standard Deviation EP Survey
Total		2,365	5,741	23,876	27,596
Dominant WQ in state of residence	1	2,277	6,880	22,550	26,331
	2	1,986	4,844	13,429	17,983
	3	2,473	5,661	25,957	27,874
	4	2,219	3,446	8,519	8,558
	5	3,075	6,600	26,444	28,643
School attendance of manager	Yes	2,626	5,928	-	-
	No	1,487	4,651	-	-
Literacy of manager	Yes	2,713	6,318	-	-
	No	1,449	2,831	-	-
Sex of manager	Male	3,448	7,459	-	-
	Female	1,518	3,327	-	-
Experience of drought in community (3y)	Yes	2,066	5,934	-	-
	No	2,458	5,658	-	-
HH ownership of mobile phone	Yes	2,489	5,663	-	-
	No	2,280	5,918	-	-
Size of locality	capital city	-	-	24,216	28,031
	over 1 mio	-	-	25,687	27,381
	250.000 to 1 mio	-	-	22,234	26,037
	50.000to250.000	-	-	27,974	29,997
	less than 50.000	-	-	32,340	26,898
highest education	no education	-	-	6966	12128
	primary completed or not	-	-	18412	22490
	secondary	-	-	17335	24142
	vocational	-	-	24770	27421
	academic	-	-	27406	25651
Any female owners	Yes	-	-	23567	27705
	No	-	-	24584	27093
Central Africa	Cameroon	-	-	28666	25576
	DRC	-	-	14809	19519
West Africa	Nigeria	-	-	12905	17822
	Gambia, The	-	-	7001	9382
	Burkina Faso	-	-	28646	27559
	Ivory Coast	-	-	10517	14582
Southern Africa	Namibia	-	-	13887	22630
	Swaziland	-	-	31097	31358
	Mozambique	-	-	13108	15881
	Zimbabwe	-	-	51892	29660
East Africa	Uganda	-	-	27928	26930
	Rwanda	-	-	40512	27298
	Kenya	-	-	31241	28303

13.11 Appendix: Descriptive Statistics of Consumption

Figure 45 shows the frequency distribution of the different levels of electricity consumption of the EP on a normal and a logarithmic scale (N=1350). On a logarithmic scale, the values are approximately normal distributed, otherwise strongly skewed to the left with a mean value of 5,682 kWh/year a median of 1712 kWh/year and a standard deviation of 12,487.[33]

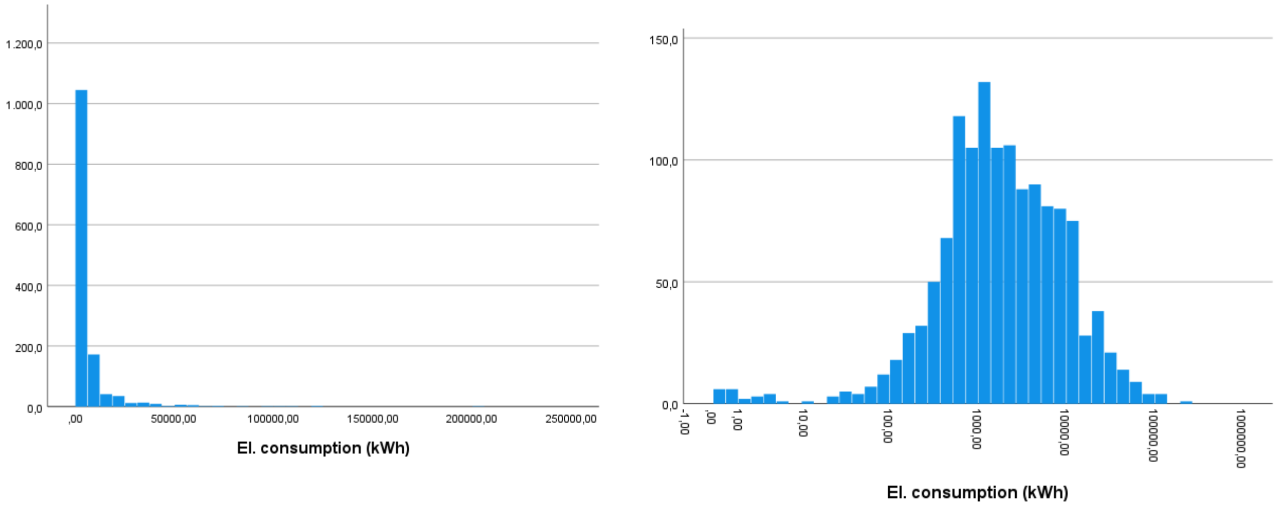

Figure 45. Frequency distribution of electricity consumption=1350.

As already suggested in section 7.3, the consumption of electricity depends on the demand of an EP due to its volume (here approximated by sales of an enterprise as seen in Figure 46) and (to a lesser degree) on the electricity supply (here in terms of outages as seen in Figure 47. As explained in the previous section, the sales naturally have their own associations with industry type, country and wealth distributions. These complex interactions and lack of qualitative insight can make it difficult to determine actual causalities and form clear narratives. Some predictors might have a weak association with consumption because they are more directly associated with sales and have a better predictive power for sales. However, Appendix 13.16 explains why the idea to test for other dependent variables such as electricity intensity (kWh/USD sales) were discarded and the choice to model the total consumption directly was deemed to be most preferable. To identify and avoid pitfalls of including predictors for outages in the model, Appendix 13.13 takes a closer look at correlations with the independent variables and outages while the present section focusses on predictors in terms of their relationship with electricity demand.

[33] Some values are above 100,000 USD because the cases were filtered for the value of sales of 100,000 USD of the survey years but for reasons of comparability, the regression was conducted for the value of USD in 2020.

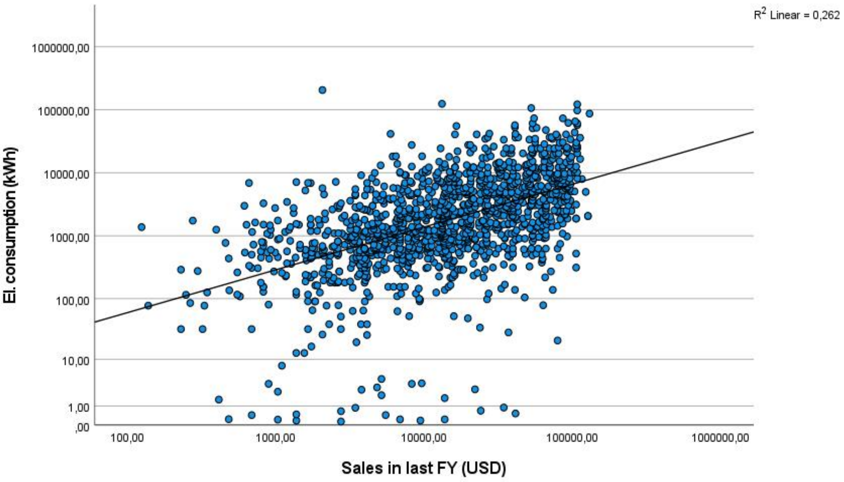

Figure 46. Scatter plot of the correlation between electricity consumption and sales on logarithmic scales.

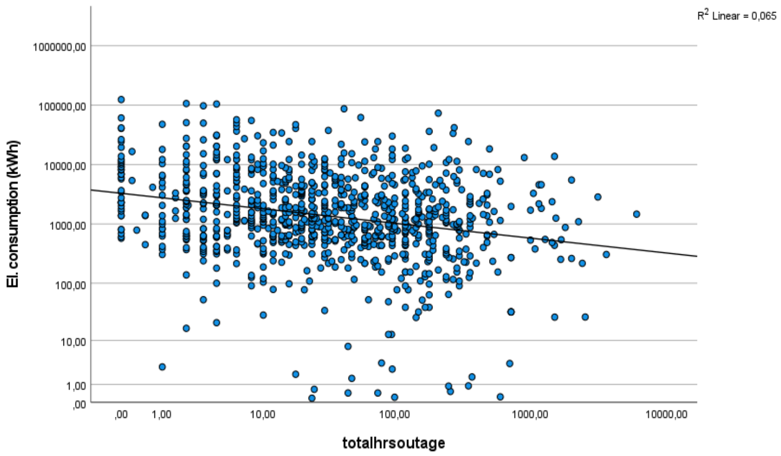

Figure 47. Relationship between electricity consumption and total number of outages experienced in a typical month.

Table 37 summarizes further findings associated with higher consumption levels. Higher education of managers, for example, has a clear tendency to operate more energy intensive businesses. Female owners tend to operate slightly less energy intensive businesses. Higher wealth tends to be associated with higher consumption with EP in regions where WQ5 is prevalent consuming more than twice as much than those where WQ2 is prevalent. An exception are EPs in areas where most people are in WQ1, which consume most electricity. To some extent, this is explained by the higher sales in the WQ, but lower efficiency might also be a possible explanation. No conclusive relationship can be found among the different locality sizes and electricity consumption. Another good predictor were the country variables, whose effects could not be explained by underlying distributions of other variables such as wealth. For the model, country-specific effect is attempted to be captured by the SSA region variable.

Appendix 13.12 illustrates the different traits of different ISICs concerning their consumption, sales, vulnerability to outages, manager characteristics, etc. for a deeper understanding of the interactions of these effects.

Table 37. Predictors of electricity consumption levels.

		El. consumption (kWh)	
		Mean	**Standard Deviation**
Highest education of manager	no education	1833	5108
	primary completed or not	1458	1948
	secondary	5101	15235
	vocational	5678	10842
	academic	5544	12805
Any female owners	yes	5419	11547
	no	5820	13073
Dominant WQ in state of residence	1	6866	14526
	2	3003	5776
	3	4671	6954
	4	5528	11444
	5	6430	13856
Size of locality	capital city	4658	8819
	over 1 mio	5238	11173
	250.000 to 1 mio	4452	12110
	50.000 to 250.000	11961	18398
	less than 50.000	1942	3323
Central Africa	Cameroon	5368	5320
	DRC	1872	3011
West Africa	Nigeria	4324	17981
	Gambia, The	819	938
	Burkina Faso	6654	10570
	Ivory Coast	2147	2634
Southern Africa	Namibia	11131	12788
	Swaziland	3705	4700
	Mozambique	10393	17624
	Zimbabwe	13554	16468
East Africa	Uganda	4409	11108
	Rwanda	3204	1483
	Kenya	3264	9347

13.12 Appendix: Description of Enterprises by ISIC

13.12.1 HH Survey (ISIC Rev4)

Retail businesses are by far the largest type of EP which is in line with literature (42%). They have little above average annual sales and are almost equally managed by men and women. The minimum age of the owner of 10 years is among the lowest. They tend to be located in rather wealthy areas. 60% are located in rural areas but they have the smallest share of agricultural engagement with only 64% of HH indicating agricultural activity. Otherwise, they have average shares of literacy (75%) and connection rates (64%). As they make up such a huge part of the sample, their characteristics also strongly dictate the total average values over all EP.

14% of EP conduct other personal service activities. Examples given by the UNSD are washing and drying activities, hairdressing, barber shops etc. Most of these businesses are run by men 62%. With a connection rate of 66% they are amongst the most frequently connected EP. They are quite evenly spread across areas with different distributions of WQ and are slightly more present in urban areas than the average (42%). Literacy is quite high at 81%.

Land transportation (defined as freight and passenger transport via rail and road), most likely to a large extend taxi services, makes up 7% of EP and is exclusively run by men, of which 89% are literate. They can be found in rather wealthy areas but have an otherwise average profile, being mostly located in rural areas and (67%). While most HH have access to electricity (59%), it is probably not related to be business itself.

Another 7% of EP engage in food and beverage service activities, i.e. cafes, bars and restaurants. 85% of them are operated by women and ownership starts at a relatively young age of 11 years. As the other women-dominated business types, the sales are well below the total average (1426 USD). Literacy amongst owners and the connection rate are slightly lower than average as well (67% and 50%). Most of the EP are situated in rural areas (75%) and areas with a slightly higher share of people in the lowest and highest WQ.

Businesses which fabricate wearing apparel make up 6% of all EP. They are mostly owned by women (68%). With only 1002 USD sales per year (less than half the amount of the average) they are the lowest earning type of EP in the sample. It is also the business with the youngest owners of only 10 years of age and lowest average age of 34 years. For being the business type with the highest share in urban areas (54% are located in urban areas), they have quite a low electricity connection rate of 55%. They tend to be located in poorer areas than the average. The literacy rate is quite high (82%).

Motorcycle trade EP make up 5% and have the highest annual sales. However, since the goods traded are rather expensive, the final profits could be much lower. 55% of owners are male. They are predominantly located in rural areas (70%) and where most people are poor. Usually, the HH has some agricultural activity as well (74%).

Food manufacturing enterprises make up 3% of the business and have a striking profile. They are almost exclusively owned by women (94%). Their sales are 35% lower than the average with on average 1545 USD/year. Only a little more than half of the owners are literate (51%). Households owning these businesses have the lowest connection rate with only 45%. This could be associated the fact that three thirds of them are located in rural areas and they have a distinct tendency to be present in rather poor areas. In 84% of cases, the HH engaged in these enterprises also have some of agricultural activity. This sector could be one of the easy-to enter survival businesses that some HH tend to engage in to avoid poverty.

Construction EP take up 2% of EP in the HH sample. They have average annual sales, are almost exclusively owned by men (and have the highest share of literate owners (97%). Despite 69% of them being located in rural areas, they have quite a high connection rate of 67%. With 19 years of age, they have the highest minimum age of owners.

13.12.2 EP Surveys (ISIC Rev3.1)

The EP Surveys have different distributions compared the Nigerian HH survey as well as differences amongst the different countries. However, some general tendencies are explained in the following.

Retail trade is again the largest category making up 37% of EP. Their electricity consumption is slightly below average with 4675 kWh/year, their annual sales of 24783 USD are almost equal to the average. They are evenly spread across WQ. Most of them are located in Southern Africa (34%). They tend to not perceive electricity access as an issue.

EP engaged in the manufacturing of wearing apparel make up almost 10% of the sample. As in the HH survey, there is a relatively high share of female owners. They consume a little less electricity than the average. However, given that their sales are 43% of the average, (lowest amongst all EP), they have the highest electricity intensity with almost 0.94 kWh per USD sales. 44% of EP are located in Central Africa. In the DRC, the Gambia and Ivory Coast samples, around 20% of business belong in this sector, all countries where EP have below average annual sales. 39% of owners did not complete secondary education, which is among the lowest rates.

8.2% of EP trade with motor vehicles. These EP have the highest share of EP without any female owners (81%). They have above average sales even though they are mostly located in less wealthy areas. Half of them are located in Eastern Africa (in the Kenya sample, 27% of EP work in this sector). They have an average annual consumption of about 5000 kWh/year but mostly do not see access to electricity as an obstacle.

Furniture manufacturing EP make up almost 6% of the sample. They are the second least energy intense business type and consume on average 3597 kWh/year. 41% of the EP are located in West Africa. Around half of them report electricity to be a major problem and 26% own a generator. Many of the owners completed their secondary education and/or had vocation training.

Making up around 6% of EP, Hotels and restaurants consume more electricity than the average but have only medium sales, i.e. are quite energy intense. 33% own a generator They tend to be present in the middle, to low WQ. Concerning education, the situation is polarized, while there are quite some owners without any formal education, there is also quite a high share with academic education

Wholesale trade EP make up 5% of the sample and are among the lowest consuming EP with 3874 kWh/year, yet most people report electricity to not be an obstacle or only a small one.

4.4% of EP manufacture food or beverages, many of them (44%) having at least one female owner. They are very energy intense and have the highest electricity consumption of all business (12000 kWh/year, twice as much as the average) and the highest sales (about 34300 USD/year, 40% more than the average). They are evenly spread across WQ and SSA regions. Half of them report electricity to be a major issue. With 14% they have the highest share of managers without any formal education.

Manufacturers of metal products have among the lowest intensity rates. Interestingly, more than half of them name electricity consumption as a major obstacle even though they are mostly located in wealthy areas.

3.2% of EP publish and print media, in some countries they make up up to 8% of EP. They have a high electricity consumption of about 7100 kWh/year. However, 80% report experiencing outages and accordingly 55% of them name electricity connection as a major obstacle. 37% own a generator. 98% have at least a secondary education and 40% even have an academic background.

Wood manufacturer make up almost 3% of EP on average but tend to be concentrated in few areas (>8% in 3 countries, Namibia and Uganda). They consume quite a lot of electricity but earn relatively little (17450 USD/year) even though they are strongly concentrated in wealthier areas. An explanation could be the low education level. 45% did not complete secondary education.

Computer and related activities businesses make up 2.4% of EP, have amongst the highest consumptions with 7000 kWh/year and are located in rather wealthy areas. They have the highest share of managers with academic background (almost 70%) and a relatively high generator ownership (42%). Maybe due to this circumstance, they rarely report electricity to be a major obstacle for operation. Perhaps, they depend so strongly on electricity that they need make sure that they have it by purchasing a generator, so it does not present an obstacle anymore.

Construction finally makes up 1.7% of businesses. They have a very low share of female ownership (18%) and the highest sales 36618 USD/Year. Having average electricity consumption, they have the lowest electricity intensity. They are mostly located in the highest WQ (53%). 38% own a generator and most of them see electricity as a rather minor or no obstacle. Owners also tend to be well educated, around a third having vocational education and almost half academic.

13.13 Appendix: Modelling Outages

To understand the extend of outages, Figure 48 gives the frequency distribution of the total number of hours of outages in a typical month experienced by the EP. As for sales and consumption, the distribution is strongly skewed to the right side with a mean value of 114, a median of 22 and a standard deviation of 392.

Investigating the connection between outages and electricity consumption further was expected to help with two issues: 1. Single out the effect of poor supply to model the full potential electricity demand of an EP and 2. Differentiate different EP types by their vulnerability to EP outages as a proxy for their electricity dependence to potentially uncover different drivers for these EP and make the model more accurate.

The large share of EP who named electricity as an obstacle for business operation and the high frequency of outages EP experienced are a strong reason to assume that most EP in the sample were not consuming at their fullest potential. around two thirds of EP (68%) indicated that they had experienced power outages in the past financial year, the interactions between consumption and outages have been captured in a conceptual model to compute the optimal consumption as dependent variable (i.e. consumption under optimal supply conditions) which will be explained in the present section. Section 0 provided evidence on the correlation of outages and consumption.

As pointed out in literature (Mayer-Tasch et al., 2013), different types of EP rely on electricity to differing degrees. While some EP seem to have a higher flexibility of consumption, others rely heavily on the availability of electricity. The concept tries to capture this dimension of dependence on electricity. A second aspect of vulnerability to outages next to the impact on consumption itself, is the impact that it has on the performance, here the sales (USD/kWh). Finally, the decision whether to start and a certain type of enterprise in a certain location will depend on these factors: flexibility of consumption, sensitivity of sales to losses in electricity consumption and the expected occurrence of outages). This means that if the information on the sensitivity of an enterprise to outages can be known, it can potentially be linked to other traits of the enterprise such as business type/industry and the information could be used, in turn, to estimate the uptake, success and electricity consumption of certain types of enterprises depending on the outages.

The information given in the surveys are the average number and length of outages experienced in a typical month in the last financial year and the estimated losses in percentage of sales due to outages (loss). Based on this, the optimal sales were calculated, I.e. sales if there had been no outages:

$$sales_{opt} = \frac{sales_{real}}{1 - loss}$$

Using the same scheme, it was attempted to calculate optimal electricity consumption:

$$cons_{opt} = \frac{cons_{real}}{1 - loss}$$

However, because of the fact that the resulting regression did not improve the models and some uncertain implicit assumptions had to be made, the idea did not end up in the final model. The main problematic assumption is that the calculation implies that the relationship between the percentage of losses in sales and the percentage of losses in consumption are the same which would mean that the relationship between sales and consumption was linear. It is more likely that there are diminishing returns per additional unit of electricity consumed or the relationship is even a stepwise function whereby a fixed amount of electricity is needed to complete a certain good or service. It is also possible that an EP has are some income generating operations which require electrification and some operations which do not rely on electricity, thus allowing for some elasticity in consumption relative to the value generated. These details are also expected to be highly dependent on the type of enterprise.

A similar idea was to calculate the losses in sales and consumption per hour of outage to better understand patterns of electricity dependence of an EP. The idea was that an EP with higher losses per hour of outage was more dependent on electricity consumption. With this information, it would have been possible to differentiate different types of enterprises, perhaps in combination with the ISIC, to then derive different consumption models. However, this again implied linearity between sales and consumption. It furthermore implies linearity between the time of operation and sales as well as time of operation and electricity consumption. Again, alternative relationships are more likely such as diminishing sales per unit of time, stepwise units of consumption per time or a more complex relationship because of other income generating activities. Furthermore, it is not clear in how far the knowledge about electricity supply might have played a role in the selection of business and reliability on electricity in the first place to there might be some selection bias. Lastly, it is neither known how consumption is spread over time nor how the outages were spread over time and therefore how much of an overlap there actually is. Some firms reported more hours of outages than hours of business operation, so they probably did not only indicate the number of outages which impacted them but estimated the total numbers of outages in a month.

Another reason that could have distorted the findings is the extent of self-generation of electricity, e.g. by means of an own generator. Furthermore, it is also possible that consumption is not hampered by the quality of supply but by the cost of consumption or the cost of appliances. But not detail was provided about these issues to they could not be attempted to be modelled.

Finally, an idea was to use the degree to which EP perceived outages as an obstacle to infer how much they depend on it in comparison with the outages they faced. Due to data availability, the findings could not be applied in the model, but it is still interesting to better understand the dependence of EP on electricity. For example, the correlation between consumption and the number of hours of outages is stronger for EP who do not perceive electricity supply as an obstacle, probably because they have a higher elasticity of consumption (Figure 49).

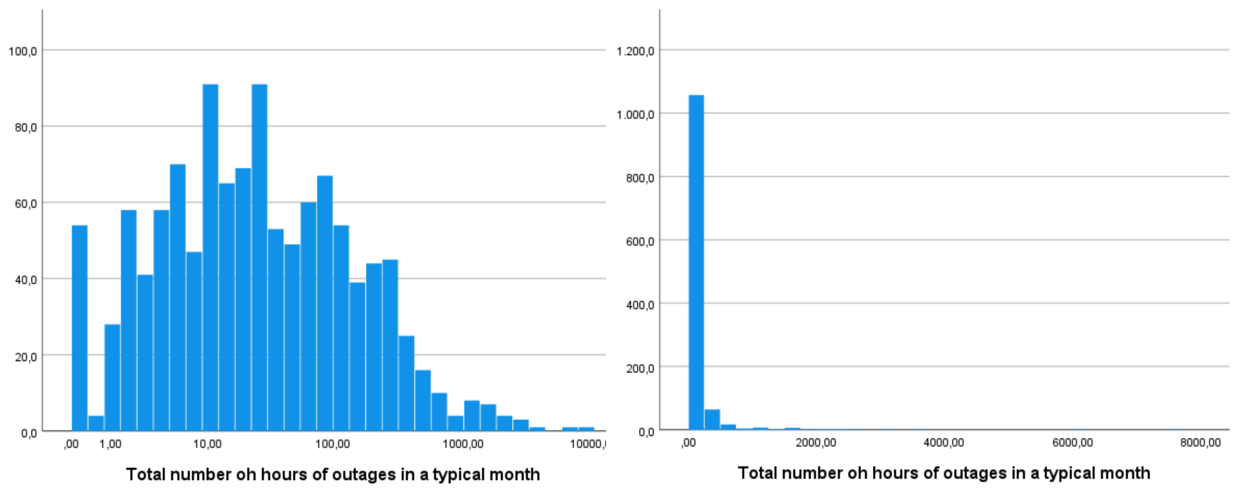

Figure 48. Frequency distribution of outages. N=1167.

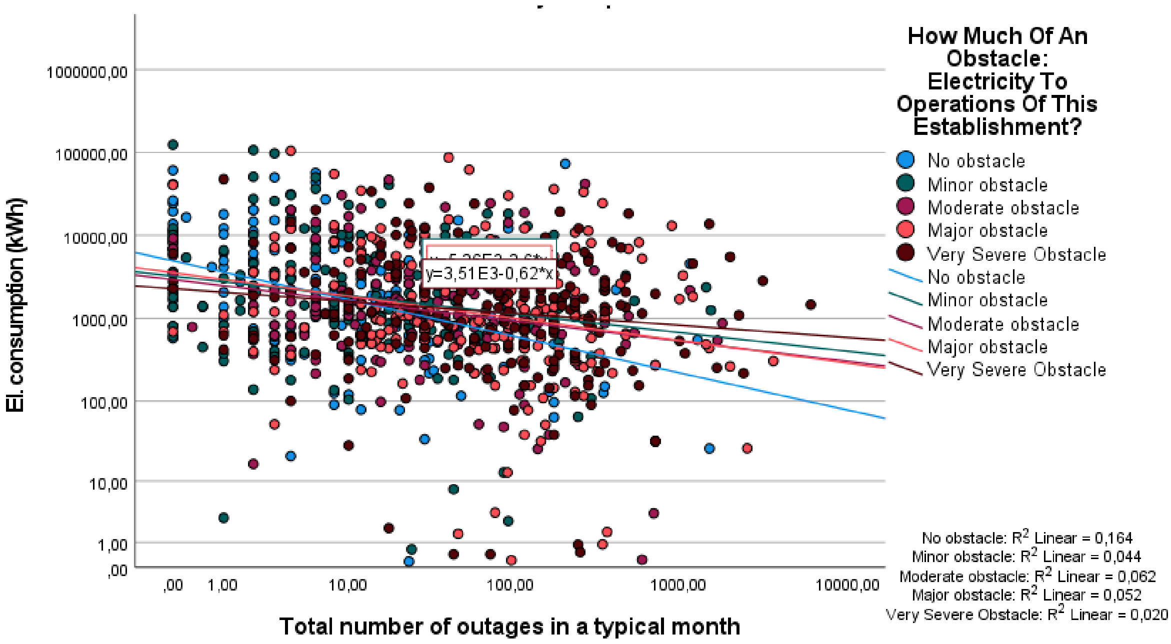

Figure 49. Relationship between electricity consumption and total number of outages experienced in a typical month depending on how far a firm perceives electricity access as an obstacle.

It was found that the more electricity is perceived as an obstacle, the less an EP consumes, i.e. they would want to consume more (Figure 49). The consumption level for those EP who do not perceive electricity as a (major) obstacle could be an indication for how much they would actually want to consume. The reasons why electricity is an obstacle are not provided but the average value of the costs of both electricity connection and consumption as well as the total number of outages are positively correlated with the degree to which EP perceive electricity as an obstacle. All findings that are in line with literature. The findings are summarized in Table 38.

Table 38. Perception of electricity as an obstacle in relationship with cost and reliability indicators.

		Mean electricity consumption (kWh)	Mean of total number of hours of outages in a typical month	Mean cost of getting an electrical connection (USD)	Mean cost of electricity as percentage of income	Mean cost of electricity (USD/kWh)
How Much Of An Obstacle: Electricity To Operations Of This Establishment?	No obstacle	6512	48	3124	18	0.20
	Minor obstacle	7292	55	4428	17	0.20
	Moderate obstacle	4866	83	5588	18	0.21
	Major obstacle	4544	156	5663	20	0.23
	Very Severe Obstacle	4095	219	5729	21	0.25

13.14 Appendix: Variables in Regression Models

Table 39. Variables in the Uptake Regression.

Category	Determinant	Available Variables	Source	Inclusion Status
Household Characteristics				
	Income/level of wealth/assets	share of population in the lowest WQ	DHS	final model
		share of population in the second WQ		final model
		share of population in the middle WQ		final model
		share of population in the fourth WQ		final model
		share of population in the highest WQ		excluded
		largest WQ (categorical)		excluded
		average WQ (scale)		excluded
		wealth index gini coefficient	DHS	excluded
	number of hh members	number of HH members	WB HH survey	final model
		avg. Number of HH members	DHS	for projection
	shock experience	occurrence of floods in the past 3 years (dummy)	WB HH survey	final model
		occurrence of droughts in the past 3 years (dummy)		final model
		occurrence of droughts	(Beguería & Vicente Serrano, 2016)	for projection
		occurrence of floods	Beguería & Vicente Serrano, 2016)	for projection
	agricultural activity (precipitation)	agricultural activity of any household member (dichotomous)	WB HH survey	final model
		share of women occupied in agriculture (m/f)	DHS	excluded
		share of men occupied in agriculture (m/f)		excluded
		share of avg. population occupied in agriculture		for projection
	access to information (e.g. radio ownership)	share of HH who own a radio	DHS	excluded
		indication if a HH owns a radio	WB HH Survey	excluded
HH Head Characteristics				
	level of education (years of schooling, literacy)	attendance of any school (dichotomous)	WB HH survey	final model
		ability to read and write in any language	WB HH survey	excluded
		share of women with secondary or higher education	DHS	excluded
		share of men with secondary or higher education		excluded
		share of avg. population with secondary or higher education		excluded
		share of women with no education	DHS	excluded
		share of men with no education		excluded
		share of avg. population with no education		excluded
		share women who attended any school	DHS	excluded
		share of men who attended any school		excluded
		share population who attended any school		for projection

Category	Determinant	Available Variables	Source	Inclusion Status
		share of literate women	DHS	excluded
		share of literate men		excluded
		share of literate avg. population		excluded
	digital and mechanical know-how	indication if a HH owns a mobile phone	WB HH Survey	final model
		share of women who use a mobile phone for financial transactions	DHS	excluded
		share of men who use a mobile phone for financial transactions		excluded
		share of overall population who use a mobile phone for financial transactions		excluded
		share of women who own a mobile phone	DHS	excluded
		share of men who own a mobile phone		excluded
		share of overall population who owns mobile phone		excluded
		share of households who own a mobile phone	DHS	for projection
		share of households who own a computer	DHS	excluded
		women who ever used the internet	DHS	excluded
		men who ever used the internet		excluded
		avg. population who ever used the internet		excluded
	age	average age of population	DHS	excluded
		share of population aged 30-49		excluded
		sex of HH head	WB HH Survey	final model
		share of female HH heads	DHS	for projection
	social network (religion)	religion	WB HH Survey	excluded
Market Access: Demand				
	population density	settlement type dummies	European Union GHSL Data Package (2019)	Final model
		Population density	WB HH Survey	Excluded
Market Access: Supply/Input factors				
	access to finance (presence of banks, interest rate, required collateral, required owners' equity contribution)	share of women with bank account	DHS	excluded
		share of men with bank account		excluded
		share of overall population with bank account		excluded
	presence electricity access	share of HH with electricity	DHS	for projection
		share of population with electricity	DHS	excluded
		nighttime lights	Elvidge et al. (2017)	final model
		availability of electricity in HH	WB HH Survey	final model
	cost of electricity consumption	national and regional electricity tariffs	different sources (see section)	excluded
Physical Market Access and Transportation				
		indication if a HH owns a television	WB HH Survey	final model

Category	Determinant	Available Variables	Source	Inclusion Status
	physical access to input goods (e.g. appliances) and services (e.g. maintenance)	share of households possessing a refrigerator	DHS	excluded
		share of households possessing a television	DHS	for projection
	level of urbanization	urban cells	Florczyk et al. (2019)	final model
		Sector (rural/urban)	WB HH Survey	final model
	distance to urban centers	distance to nearest population center (>20.000 inhabitants) (km)	WB HH survey/WorldCities	final model
		travel time to the next city	Weiss et al. (2018)	final model
	presence of roads and distance to roads	distance to nearest major road (km)	WB HH survey/FERMA	final model
		road density	CIESIN et al. (2013)	excluded
	presence of regular market/commercial center	distance to nearest key market centers (km)	WB HH survey/USAID FEWSNET	final model
Governance, institutional context and development				
		Human Development Index	Global Data Lab	excluded/highly correlated
	effort to start a business	national/regional Distance to Frontier score of the Doing Business Indicator for the ease of starting a business	WB Doing Business	final model
		The rank of ease of starting a business	WB Doing Business	excluded
		The expected number of days it takes to start a business	WB Doing Business	excluded
		The expected number of procedures it takes to start a business	WB Doing Business	excluded
		The expected cost required to start a business	WB Doing Business	excluded
		The expected capital required to start a business	WB Doing Business	excluded
	weather (droughts)	precipitation of the wettest month	WB HH Survey	excluded
		annual mean temperature	WB HH Survey	excluded
		annual precipitation	WB HH Survey	excluded
		mean temperature of the wettest month	WB HH Survey	excluded
Further potential predictors				
	(alternative) employment opportunities	share of women who worked in the last 12 months and are currently	DHS	final model
		share of men who worked in the last 12 months and are currently		final model
		share of avg. population who worked in the last 12 months and are currently		excluded

Table 40. Variables tested for the connection regression.

Category	Determinant	Available Variables	Source	Inclusion Status
Household Characteristics				
	Income/level of wealth/assets	share of population in the lowest WQ	DHS	excluded
		share of population in the second WQ		excluded
		share of population in the middle WQ		excluded

Category	Determinant	Available Variables	Source	Inclusion Status
		share of population in the fourth WQ		excluded
		share of population in the highest WQ		final model
		largest WQ (categorical)		excluded
		average WQ (scale)		excluded
		wealth index gini coefficient	DHS	excluded
	number of hh members	number of HH members	WB HH survey	excluded
		avg. Number of HH members	DHS	excluded
	shock experience	occurrence of floods in the past 3 years (dummy)	WB HH survey	excluded
		occurrence of droughts in the past 3 years (dummy)		final model
		occurrence of droughts	yes	for projection
		occurrence of floods	yes	excluded
	agricultural activity (precipitation)	agricultural activity of any household member (dichotomous)	WB HH survey	final model
		share of women occupied in agriculture (m/f)	DHS	excluded
		share of men occupied in agriculture (m/f)		excluded
		share of avg. population occupied in agriculture		for projection
Entrepreneur Characteristics				
	level of education (years of schooling, literacy)	attendance of any school (dichotomous)	WB HH survey	excluded
		ability to read and write in any language	WB HH survey	final model
		share of women with secondary or higher education	DHS	excluded
		share of men with secondary or higher education		excluded
		share of avg. population with secondary or higher education		excluded
		share of women with no education		excluded
		share of men with no education		excluded
		share of avg. population with no education		excluded
		share women who attended any school	DHS	excluded
		share of men who attended any school		excluded
		share population who attended any school		excluded
		share of literate women	DHS	excluded
		share of literate men		excluded
		share of literate avg. population		excluded
	digital and mechanical know-how	indication if a HH owns a mobile phone	WB HH Survey	final model
		share of women who use a mobile phone for financial transactions	DHS	excluded
		share of men who use a mobile phone for financial transactions		excluded
		share of overall population who use a mobile phone for financial transactions		excluded

Category	Determinant	Available Variables	Source	Inclusion Status
		share of women who own a mobile phone	DHS	excluded
		share of men who own a mobile phone		excluded
		share of overall population who owns mobile phone		excluded
		share of households who own a mobile phone	DHS	for projection
		share of households who own a computer	DHS	excluded
	social networks	women who ever used the internet	DHS	excluded
		men who ever used the internet		excluded
		avg. population who ever used the internet		excluded
	age	average age of population	DHS	excluded
		share of population aged 30-49		excluded
	sex	sex of manager	WB HH Survey	excluded
		sex of HH head	WB HH Survey	excluded
		share of female HH heads	DHS	excluded
	religion	religion	WB HH Survey	excluded
Enterprise Characteristics				
	wealth (income, assets, capital)	natural logarithm of total sales of the last fiscal year of the enterprise (USD 2020)	WB HH Surveys	final model
	type/industry	2-digit ISIC 4	WB HH Survey	excluded
Market Access: Demand				
	population density	aera type dummies	European Union GHSL Data Package (2019)	excluded
		Population density	WB HH Survey	excluded
Market Access: Supply/Input factors				
	physical access to input goods (e.g. appliances) and services (e.g. maintenance)	indication if a HH owns a television	WB HH Survey	excluded
		share of households possessing a refrigerator	DHS	excluded
		share of households possessing a television	DHS	excluded
	access to finance (presence of banks, interest rate, required collateral, required owners' equity contribution)	share of women with bank account	DHS	excluded
		share of men with bank account		excluded
		share of overall population with bank account		excluded
	cost of electricity consumption	national and regional electricity tariffs	different sources (see section)	excluded
Physical Market Access and Transportation				
	level of urbanization	urban cells	Florczyk et al. (2019)	final model
		Sector (rural/urban)	WB HH Survey	excluded
	distance to urban centers	distance to nearest population center (>20.000 inhabitants) (km)	WB HH survey/WorldCities	excluded
		travel time to the next city	Weiss et al. (2018)	excluded
	presence of roads and distance to roads	distance to nearest major road (km)	WB HH survey/FERMA	excluded
		road density	CIESIN et al. (2013)	excluded

Category	Determinant	Available Variables	Source	Inclusion Status
	presence of regular market/commercial center	distance to nearest key market centers (km)	WB HH survey/USAID FEWSNET	excluded
Governance, institutional context and development				
		Human Development Index	Global Data Lab	excluded
	effort to start a business	national/regional Distance to Frontier score of the Doing Business Indicator for the ease of starting a business	WB Doing Business	excluded
		The rank of ease of starting a business	WB Doing Business	excluded
		The expected number of days it takes to start a business	WB Doing Business	excluded
		The expected number of procedures it takes to start a business	WB Doing Business	excluded
		The expected cost required to start a business	WB Doing Business	excluded
		The expected capital required to start a business	WB Doing Business	excluded
	weather (droughts)	precipitation of the wettest month	WB HH Survey	excluded
		annual mean temperature	WB HH Survey	excluded
		annual precipitation	WB HH Survey	excluded
		mean temperature of the wettest month	WB HH Survey	excluded
Further Potential Predictors				
	(alternative) employment opportunities	share of women who worked in the last 12 months and are currently	DHS	excluded
		share of men who worked in the last 12 months and are currently		excluded
		share of avg. population who worked in the last 12 months and are currently		excluded

Table 41. Variables testes in the sales model.

Category	Determinant	Variables available	Source	Inclusion
Household Characteristics				
	Income/level of wealth/assets	share of population in the lowest WQ	DHS	final model
		share of population in the second WQ		final model
		share of population in the middle WQ		final model
		share of population in the fourth WQ		final model
		share of population in the highest WQ		excluded
		largest WQ (categorical)		excluded
		average WQ (scale)		excluded
		wealth index gini coefficient	DHS	excluded
	agricultural activity (precipitation)	share of women occupied in agriculture (m/f)	DHS	excluded
		share of men occupied in agriculture (m/f)		excluded
		share of avg. population occupied in agriculture		excluded
Entrepreneur Characteristics				
	level of education (years of schooling, literacy)	share of women with secondary or higher education	DHS	excluded
		share of men with secondary or higher education		excluded
		share of avg. population with secondary or higher education		excluded
		share of women with no education		excluded

Category	Determinant	Variables available	Source	Inclusion
		share of men with no education		excluded
		share of avg. population with no education		final model
		share women who attended any school		excluded
		share of men who attended any school		excluded
		share population who attended any school		excluded
		share of literate women		excluded
		share of literate men		excluded
		share of literate avg. population		excluded
	digital and mechanical know-how	share of women who use a mobile phone for financial transactions	DHS	excluded
		share of men who use a mobile phone for financial transactions		excluded
		share of overall population who use a mobile phone for financial transactions		excluded
		share of women who own a mobile phone	DHS	excluded
		share of men who own a mobile phone		excluded
		share of overall population who owns mobile phone		excluded
		share of households who own a mobile phone	DHS	excluded
		share of households who own a computer	DHS	excluded
		women who ever used the internet	DHS	excluded
		men who ever used the internet		
		avg. population who ever used the internet		
	age	average age of population	DHS	excluded
		share of population aged 30-49		excluded
	sex	share of female HH heads	DHS	final model
Enterprise Characteristics				
	type/industry	3-digit ISIC 3.1	WB EP Survey	excluded
		2-digit ISIC 3.1		excluded
		categorical ISIC 3.1		excluded
Market Access (general)				
Market Access: Demand				
Market Access: Supply/Input factors				
	access to finance (presence of banks, interest rate, required collateral, required owners' equity contribution)	share of women with bank account	DHS	excluded
		share of men with bank account		excluded
		share of overall population with bank account		excluded
		Firms using banks to finance investment (% of firms)	WB Development Indicators	excluded
		Firms using banks to finance working capital (% of firms)	WB Development Indicators	excluded
		Lending interest rate (%)	WB Development Indicators	final model
	presence electricity access	share of HH with electricity	DHS	excluded
		share of population with electricity	DHS	excluded
		Access to electricity, urban (% of urban population)	WB Development Indicators	excluded
		Access to electricity, rural (% of rural population)	WB Development Indicators	excluded
		Access to electricity (% of population)	WB Development Indicators	excluded
		regional score of the Doing Business Indicator for getting an electrical connection	WB Doing Business	final model
	electricity access capacity (voltage, AC/DC, etc.) quality (number and duration of	Power outages in firms in a typical month (number)	WB PE Survey/ Development Indicators	final model

Category	Determinant	Variables available	Source	Inclusion
	blackouts in given time; presence of risk factors for supply)	Value lost due to electrical outages (% of sales for affected firms)	WB Development Indicators	excluded
		Firms experiencing electrical outages (% of firms)	WB Development Indicators	excluded
		Electric power transmission and distribution losses (% of output)	WB Development Indicators	excluded
		experience of outages in last FY (dichotomous)	WB EP Surveys	excluded
	cost of electrical connection	getting electricity cost	WB Doing Business	excluded
	cost of electricity consumption	national and regional electricity tariffs	different sources (see section)	excluded
Physical Market Access and Transportation				
	physical access to input goods (e.g. appliances) and services (e.g. maintenance)	share of households possessing a refrigerator	DHS	excluded
		share of households possessing a television	DHS	excluded
Governance, institutional context and development				
	effort to start a business	national/regional Distance to Frontier score of the Doing Business Indicator for the ease of starting a business	WB Doing Business	excluded
		The rank of ease of starting a business	WB Doing Business	excluded
		The expected number of days it takes to start a business	WB Doing Business	excluded
		The expected number of procedures it takes to start a business	WB Doing Business	excluded
		The expected cost required to start a business	WB Doing Business	excluded
		The expected capital required to start a business	WB Doing Business	excluded
	country	SSA region dummy	WB EP Survey	final model
	GDP	GDP PPP	WB Development Indicators	excluded
		GDP PPP/capita	WB Development Indicators	excluded
		GNI PPP	WB Development Indicators	excluded
		GNI PPP/capita	WB Development Indicators	excluded
Further potential predictors				
	availability of digital infrastructure and digital know-how	Fixed broadband subscriptions (per 100 people)	WB Development Indicators	excluded
		Mobile cellular subscriptions (per 100 people)	WB Development Indicators	excluded
	(alternative) employment opportunities	share of population who worked in the last 12 months and are currently	DHS	excluded
		Vulnerable employment, total (% of total employment) (modeled ILO estimate)	WB Development Indicators	excluded
	Prevalence of poverty	Poverty headcount ratio at $1.90 a day ((2011) PPP) (% of population)	WB Development Indicators	excluded
		Population living in slums (% of urban population)	WB Development Indicators	excluded
	Other development indicators	Human Development Index	Global Data Lab	excluded
		Net ODA received (% of GNI)	WB Development Indicators	excluded

Table 42. Variables tested in the consumption regression.

Category	Determinant	Variables available	Source	Inclusion
Household Characteristics				
	Income/level of wealth/assets	share of population in the lowest WQ	DHS	excluded
		share of population in the second WQ		final model
		share of population in the middle WQ		final model
		share of population in the fourth WQ		final model
		share of population in the highest WQ		final model
		largest WQ (categorical)		excluded
		average WQ (scale)		excluded
		wealth index gini coefficient	DHS	excluded
	agricultural activity (precipitation)	share of women occupied in agriculture (m/f)	DHS	excluded
		share of men occupied in agriculture (m/f)		excluded
		share of avg. population occupied in agriculture		excluded
Entrepreneur Characteristics				
	level of education (years of schooling, literacy)	share of women with secondary or higher education	DHS	excluded
		share of men with secondary or higher education		excluded
		share of avg. population with secondary or higher education		excluded
		share of women with no education		excluded
		share of men with no education		excluded
		share of avg. population with no education		excluded
		share women who attended any school		excluded
		share of men who attended any school		excluded
		share population who attended any schol		excluded
		share of literate women		excluded
		share of literate men		excluded
		share of literate avg. population		excluded
	digital and mechanical know-how	share of women who use a mobile phone for financial transactions	DHS	excluded
		share of men who use a mobile phone for financial transactions		excluded
		share of overall population who use a mobile phone for financial transactions		excluded
		share of women who own a mobile phone	DHS	excluded
		share of men who own a mobile phone		excluded
		share of overall population who owns mobile phone		excluded
		share of households who own a mobile phone	DHS	excluded
		share of households who own a computer	DHS	excluded
		women who ever used the internet	DHS	excluded
		men who ever used the internet		excluded
		avg. population who ever used the internet		excluded
	age	average age of population	DHS	excluded
		share of population aged 30-49	DHS	excluded
	sex	share of female HH heeds	DHS	final model
Enterprise Characteristics				
	wealth (income, assets, capital)	natural logarithm of total sales of the last fiscal year of the enterprise (USD 2020)	WB EP Surveys	final model

Category	Determinant	Variables available	Source	Inclusion
	type/industry	3-digit ISIC 3.1	WB EP Survey	excluded
		2-digit ISIC 3.1		excluded
		categorical ISIC 3.1		excluded
Market Access (general)				
Market Access: Demand				
Market Access: Supply/Input factors				
	access to finance (presence of banks, interest rate, required collateral, required owners' equity contribution)	share of women with bank account	DHS	excluded
		share of men with bank account		excluded
		share of overall population with bank account		excluded
		Firms using banks to finance investment (% of firms)	WB Development Indicators	excluded
		Firms using banks to finance working capital (% of firms)	WB Development Indicators	excluded
		Lending interest rate (%)	WB Development Indicators	excluded
	presence electricity access	share of HH with electricity	DHS	excluded
		share of population with electricity	DHS	excluded
		regional score of the Doing Business Indicator for getting an electrical connection	WB Doing Business	final model
	electricity access capacity (voltage, AC/DC, etc.) quality (number and duration of blackouts in given time; presence of risk factors for supply)	Power outages in firms in a typical month (number)	WB Development Indicators	excluded
		Value lost due to electrical outages (% of sales for affected firms)	WB Development Indicators	excluded
		Firms experiencing electrical outages (% of firms)	WB Development Indicators	excluded
		Electric power transmission and distribution losses (% of output)	WB Development Indicators	excluded
		experience of outages in last FY (dichotomous)	WB EP Surveys	excluded
		number of outages in an average month in last FY	WB EP Surveys/WB Development Indicators	final model
	cost of electrical connection	getting electricity cost	WB Doing Business	excluded
	cost of electricity consumption	national and regional electricity tariffs	different sources (see section)	excluded
Governance, institutional context and development				
	effort to start a business	national/regional Distance to Frontier score of the Doing Business Indicator for the ease of starting a business	WB Doing Business	final model
		The rank of ease of starting a business	WB Doing Business	excluded
		The expected number of days it takes to start a business	WB Doing Business	excluded
		The expected number of procedures it takes to start a business	WB Doing Business	excluded
		The expected cost required to start a business	WB Doing Business	excluded
		The expected capital required to start a business	WB Doing Business	excluded
	country	SSA region dummy	WB EP Survey	final model

Category	Determinant	Variables available	Source	Inclusion
	GDP	GDP PPP	WB Development Indicators	excluded
		GDP PPP/capita	WB Development Indicators	excluded
		GNI PPP	WB Development Indicators	final model
		GNI PPP/capita	WB Development Indicators	excluded
	access to water	access to water and quality of supply	WB EP survey	excluded
Further potential predictors				
	availability of digital infrastructure and digital know-how	Fixed broadband subscriptions (per 100 people)	WB Development Indicators	excluded
		Mobile cellular subscriptions (per 100 people)	WB Development Indicators	excluded
	(alternative) employment opportunities	share of population who worked in the last 12 months and are currently	DHS	excluded
		Vulnerable employment, total (% of total employment) (modeled ILO estimate)	WB Development Indicators	excluded
	Prevalence of poverty	Poverty headcount ratio at $1.90 a day ((2011) PPP) (% of population)	WB Development Indicators	excluded
		Population living in slums (% of urban population)	WB Development Indicators	excluded
	Other development indicators	Human Development Index	Global Data Lab	excluded
		Net ODA received (% of GNI)	WB Development Indicators	excluded

13.15 Appendix: Testing of correlations between predictors and dependent variables

Table 43. Testing for significant differences between EP HH and non-EP HH by column proportions (z-test) and column means (t-test) of the given predictors.

		HH without EP		HH with EP	
		Count	**Mean**	**Count**	**Mean**
Access to electricity in HH	Yes	5143849$_a$		9802240$_b$	
	No	5763643$_a$		6247320$_b$	
Share of women employed in state			69.39$_a$		65.00$_b$
Share of men employed in state			84.70$_a$		85.30$_b$
Agricultural activity in HH	Yes	8266940$_a$		10670883$_b$	
	No	2640552$_a$		5378677$_b$	
Presence partner of HHH	Yes	7458580$_a$		12748843$_b$	
	No	3448912$_a$		3300717$_b$	
Sex of HHH	Male	8540109$_a$		13394168$_b$	
	Female	2367383$_a$		2655392$_b$	
School attendance of HHH	Yes	7689370$_a$		13417313$_b$	
	No	3218122$_a$		2632247$_b$	
Number of HH members			5.70$_a$		6.50$_b$
Sector	Urban	2475480$_a$		5999445$_b$	
	Rural	8432012$_a$		10050116$_b$	
Population density			1599$_a$		2300$_b$
Number of urban cells			9.22$_a$		9.22$_b$
Travel time to the next city			9.22$_a$		9.22$_b$
Nighttime lights			107.33$_a$		103.03$_b$
Area type	water grid cell	38276$_a$		25517$_b$	
	very low-density rural grid cell	8914595$_a$		10770797$_b$	
	low density rural grid cell	156578$_a$		221511$_b$	
	rural cluster grid cell	54968$_a$		113607$_b$	
	suburban or peri-urban grid cell	87694$_a$		143107$_b$	
	semi-dense urban cluster grid cell	0[5]		0[5]	
	dense urban cluster grid cell	0[5]		0[5]	
	urban center grid cell	1655382$_a$		4775021$_b$	
Share of pop. in lowest WQ in state			13.99$_a$		17.04$_b$
Share of pop. in low-mid WQ in state			18.78$_a$		17.56$_b$
Share of pop. in middle WQ in state			22.17$_a$		20.63$_b$
Share of pop. in mid-hi WQ in state			22.95$_a$		22.52$_b$
Share of pop. in highest WQ in state			22.13$_a$		22.25$_b$
Dominating WQ in state	1	1166725$_a$		3603393$_b$	
	2	2687592$_a$		3110870$_b$	
	3	2814477$_a$		3242720$_b$	
	4	1635559$_a$		2349553$_b$	
	5	2603138$_a$		3743024$_b$	
Distance to nearest major road			4.92$_a$		5.30$_b$
Distance to nearest population center			22.10$_a$		22.61$_b$
HH ownership of mobile phone	Yes	6666922$_a$		10953713$_b$	
	No	4226967$_a$		5095847$_b$	
HH ownership of TVs	Yes	4295802$_a$		7865060$_b$	

		HH without EP		HH with EP	
		Count	Mean	Count	Mean
	No	6610175_a		8184500_b	
Experience of drought in community (3y)	Yes	1899838_a		2502557_b	
	No	8567112_a		13009678_b	
Experience of flood in community (3y)	Yes	4332309_a		5960835_b	
	No	6134641_a		9551399_b	
Share of population with access to electricity			58.35_a		59.91_b
Share of population aged 30-49 in state			21.11_a		20.64_b
HDI			$.561_a$		$.547_b$
Share of HH with access to electricity in state			58.50_a		59.20_b
Annual Precipitation (mm)			1563_a		1440_b
Precipitation n of Wettest Quarter			759_a		726_b
share of hh who own a computer			22.77_a		26.92_b
share of population without any formal education			28.21_a		32.88_b
share of women without any formal education			17.34_a		20.97_b
share of men without any formal education			23.46_a		23.25_b
average age in state			4.53_a		4.72_b
average number of hh members			6.28_a		6.25_b
share of hh who own a TV			50.89_a		48.94_b
Precipitation of Wettest Month (mm)			294_a		286_b
share of population occupied in agriculture			33.57_a		27.97_b
Starting a Business: number of procedures			9.83_a		9.80_b
Starting a Business: number of days			25.81_a		24.71_b
Starting a Business: cost			2887121289.73_a		2782550634.33_b
Starting a Business: rank			18.81_a		17.18_b
roads density			9869.92_a		12633.07_b
average share of people who are employed			77.05_a		75.15_b
share of female hhh			20.34_a		18.31_b
average number of hh members			4.47_a		4.67_b
share of population who ever used the internet			28.96_a		28.97_a
share of women with secondary education			12.12_a		11.65_b
Starting a Business: capital			.00[1]		.00[1]
Annual Mean Temperature (degC * 10)			263_a		263_b
Mean Temperature of Wettest Quarter (degC * 10)			251_a		252_b
share of population with a bank account			32.99_a		31.65_b
share of population with a mobile phone			70.29_a		69.74_b
share of population who uses a phone for banking			29.23_a		28.42_b
share of hh who own a mobile phone			88.80_a		87.65_b
share of hh who own a fridge			22.19_a		21.97_b
share of literate women			59.29_a		56.93_b
share of men with secondary education			17.19_a		16.55_b
share of literate men			76.53_a		73.38_b
share of women who ever used the internet			18.32_a		18.04_b
share of men who ever used the internet			39.60_a		39.90_b

		HH without EP		HH with EP	
		Count	Mean	Count	Mean
HH ownership of bikes	Yes	1493514_a		2741875_b	
	No	9412463_a		13304142_b	
HH ownership of motorbikes	Yes	2856039_a		5380680_b	
	No	8049938_a		10668880_b	
HH ownership of cars and other vehicles	Yes	967982_a		1631609_b	
	No	9937995_a		14402276_b	
HH ownership of radio	Yes	255316_a		435210_b	
	No	10650661_a		15610807_b	
HH ownership of computer	Yes	436923_a		638069_b	
	No	10469054_a		15392718_b	

Note: Values in the same row and sub-table not sharing the same subscript are significantly different at p< .05 in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. Tests assume equal variances.[2,6]

1. This category is not used in comparisons because there are no other valid categories to compare

2. Tests are adjusted for all pairwise comparisons within a row of each innermost sub-table using the Bonferroni correction.

5. This category is not used in comparisons because its column proportion is equal to zero or one.

6. Cell counts of some categories are not integers. They were rounded to the nearest integers before performing column proportions tests.

Table 44. Testing for significant differences between EP with and without electricity access by column proportions (z-test) and column means (t-test) of the given predictors.

		EP without electricity		EP with electricity	
		Count	Mean	Count	Mean
Literacy of manager	Yes	979_a		2046_b	
	No	506_a		382_b	
Agricultural activity in HH	Yes	1299_a		1304_b	
	No	186_a		1124_b	
HH ownership of mobile phone	Yes	894_a		1845_b	
	No	591_a		583_b	
Experience of drought in community (3y)	Yes	447_a		230_b	
	No	995_a		2101_b	
Number of urban cells			9.18_a		9.22_b
Sector	Urban	207_a		1265_b	
	Rural	1278_a		1163_b	
Population density			740_a		3933_b
Number of HH members			7.76_a		6.60_b
Distance to nearest major road			4.83_a		4.82_a
Distance to nearest population center			22.06_a		21.91_a
Distance to nearest market			68.40_a		68.39_a
HDI			$.499_a$		$.581_b$
School attendance of manager	Yes	1075_a		2153_b	
	No	410_a		275_b	
Sex of manager	Male	722_a		1119_a	
	Female	763_a		1309_a	
Experience of flood in community (3y)	Yes	663_a		759_b	
	No	779_a		1572_b	
share of pop. in lowest WQ in state			25.25_a		10.75_b
share of pop. in low-mid WQ in state			22.82_a		14.58_b
share of pop. in mid WQ in state			20.92_a		20.82_a
share of pop. in mid-hi WQ in state			17.56_a		25.27_b

		EP without electricity		EP with electricity	
		Count	Mean	Count	Mean
share of HH in hi WQ in state			13.44_a		28.58_b
Dominating WQ in state	1	542_a		300_b	
	2	384_a		390_b	
	3	303_a		468_a	
	4	137_a		516_b	
	5	119_a		754_b	
Travel time to the next city			9.22_a		9.22_b
Nighttime lights			25.55_a		162.68_b
Road Density			10390.73_a		13023.26_b
Starting a Business DtF			77.16_a		77.74_b
Annual Mean Temperature (degC * 10)			263_a		263_a
Mean Temperature of Wettest Quarter (degC * 10)			253_a		251_b
Annual Precipitation (mm)			1170_a		1600_b
Precipitation of Wettest Month (mm)			264_a		303_b
Precipitation of Wettest Quarter			658_a		773_b
Share of women employed in state			60.43_a		67.40_b
Share of men employed in state			87.23_a		82.99_b
Area type	water grid cell	6_a		0[2]	
	very low-density rural grid cell	1307_a		1380_b	
	low density rural grid cell	24_a		21_b	
	rural cluster grid cell	33_a		19_b	
	suburban or peri-urban grid cell	4_a		22_b	
	semi-dense urban cluster grid cell	0[2]		0[2]	
	dense urban cluster grid cell	0[2]		0[2]	
	urban center grid cell	111_a		986_b	
HH ownership of TV		248_a		1764_b	
Share of population with access to electricity			49.91_a		68.44_b
Share of population aged 30-49 in state			18.67_a		22.01_b
Share of HH with access to electricity in state			48.82_a		68.11_b
share of hh who own a computer			5.18_a		7.35_b
share of population without any formal education			37.40_a		18.65_b
share of women without any formal education			45.72_a		23.22_b
share of men without any formal education			29.09_a		14.07_b
average age in state			22.00_a		24.02_b
average number of hh members			5.20_a		4.39_b
share of hh who own a TV			36.75_a		57.36_b
Precipitation of Wettest Month (mm)			264_a		303_b
Experience of flood in community in past 3 years	Yes	135_a		87_b	
	No	1350_a		2341_b	
share of population occupied in agriculture			31.75_a		24.71_b
Starting a Business: number of procedures			9.76_a		9.67_b
Starting a Business: number of days			25.58_a		23.76_b
Starting a Business: cost			2834445362.80_a		2914776274.27_b
Starting a Business: rank			17.55_a		16.92_a
average share of people who are employed			73.83_a		75.19_b
share of female hhh			14.32_a		20.56_b

		EP without electricity		EP with electricity	
		Count	Mean	Count	Mean
average number of hh members			5.16_{a}		4.38_{b}
share of population who ever used the internet			22.58_{a}		32.94_{b}
share of women with secondary education			8.44_{a}		14.18_{b}
share of population with a bank account			22.46_{a}		37.18_{b}
share of population with a mobile phone			62.84_{a}		74.42_{b}
share of population who uses a phone for banking			20.79_{a}		32.92_{b}
share of hh who own a mobile phone			83.98_{a}		89.96_{b}
share of hh who own a fridge			15.26_{a}		27.00_{b}
share of literate women			44.45_{a}		64.97_{b}
share of men with secondary education			15.28_{a}		18.15_{b}
share of literate men			65.96_{a}		77.83_{b}
share of women who ever used the internet			11.40_{a}		22.93_{b}
share of men who ever used the internet			33.76_{a}		42.96_{b}
HH ownership of bikes	Yes	1493514_{a}		2741875_{b}	
	No	9412463_{a}		13304142_{b}	
HH ownership of motorbikes	Yes	2856039_{a}		5380680_{b}	
	No	8049938_{a}		10668880_{b}	
HH ownership of cars and other vehicles	Yes	967982_{a}		1631609_{b}	
	No	9937995_{a}		14402276_{b}	
HH ownership of radio	Yes	255316_{a}		435210_{b}	
	No	10650661_{a}		15610807_{b}	
HH ownership of computer	Yes	436923_{a}		638069_{b}	
	No	10469054_{a}		15392718_{b}	

Note: Values in the same row and sub-table not sharing the same subscript are significantly different at p< .05 in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. Tests assume equal variances.[1]

1. Tests are adjusted for all pairwise comparisons within a row of each innermost sub-table using the Bonferroni correction.

2. This category is not used in comparisons because its column proportion is equal to zero or one.

Table 45. Bivariate Pearson correlations between the independent variables. outages. electricity consumption and sales.

	Number of outages in a typical month			ln El. consumption last FY (kWh)			ln Sales in last FY (USD)		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Number of outages in a typical month	1		2262				-0.068	0.003	1960
Natural logarithm of total sales in the last financial year	-0.068	0.003	1960	0.083	0.002	1339	1		1960
Share of female HHH in state	-0.055	0.014	2026	-0.153	0.000	1332	0.249	0.000	1796
Share of women with secondary education in state	-0.048	0.025	2143	0.157	0.000	1402	-0.157	0.000	1850
GNI. PPP (constant 2017 international $)	-0.040	0.058	2262	0.228	0.000	1438	-0.08	0.000	1960
National net ODA received (% of GNI)	-0.008	0.701	2262	0.178	0.000	1438	-0.192	0.000	1960
Doing Business: Starting a Business Dtf score	0.048	0.022	2262	0.434	0.000	1438	-0.284	0.000	1960

	Number of outages in a typical month			ln El. consumption last FY (kWh)			ln Sales in last FY (USD)		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Doing Business: Getting Electricity score	-0.114	0.000	2262	0.133	0.000	1438	0.09	0.000	1960
Share of pop. without education	0.004	0.843	2167	0.024	0.369	1409	-0.266	0.000	1874
National lending interest rate (%)	-0.136	0.000	2167	0.399	0.000	1408	-0.044	0.056	1866
share of pop. in lowest WQ in state	-0.014	0.535	2026	0.034	0.210	1332	-0.086	0.000	1796
share of pop. in low-mid WQ in state	-0.014	0.527	2026	0.040	0.149	1332	-0.152	0.000	1796
share of pop. in middle WQ in state	0.08	0.000	2026	0.053	0.053	1332	-0.198	0.000	1796
share of pop. in mid-hi WQ in state	0.046	0.037	2026	-0.198	0.000	1332	-0.020	0.407	1796
share of pop. in highest WQ in state	-0.029	0.189	2026	0.019	0.487	1332	0.179	0.000	1796
share of literate women	-0.023	0.280	2143	-0.106	0.000	1402	0.313	0.000	1850
share of men with secondary education	-0.094	0.000	1749	0.24	0.000	1224	-0.194	0.000	1502
share of literate men	-0.020	0.403	1749	-0.12	0.000	1224	0.294	0.000	1502
share of HH with access to electricity	0.042	0.062	2026	-0.003	0.909	1332	0.073	0.002	1796
share of population with access to electricity	0.047	0.036	2026	0.011	0.686	1332	0.043	0.066	1796
share of HH who own a TV	0.042	0.062	2026	-0.024	0.383	1332	0.043	0.066	1796
share of HH who own a mobile phone	0.047	0.035	2026	-0.163	0.000	1332	0.1	0.000	1796
share of HH who own a computer	-0.041	0.100	1624	-0.156	0.000	1076	0.296	0.000	1395
share of HH who own a fridge	-0.004	0.872	2026	-0.138	0.000	1332	0.029	0.225	1796
average wealth score per area	-0.006	0.779	2026	-0.025	0.368	1332	0.146	0.000	1796
share of men who ever used the internet	-0.050	0.097	1090	-0.286	0.000	786	0.458	0.000	949
share of men who use a mobile phone for banking	-0.050	0.097	1090	-0.286	0.000	786	0.458	0.000	949
share of men who own a mobile phone	0.091	0.003	1090	0.075	0.035	786	0.079	0.015	949
share of men who have a bank account	0.156	0.000	1090	0.142	0.000	786	-0.146	0.000	949
share of women who use a mobile phone for banking	-0.032	0.297	1090	-0.322	0.000	786	0.482	0.000	949
share of women who own a mobile phone	0.033	0.276	1090	-0.167	0.000	786	0.336	0.000	949
share of women who have a bank account	0.173	0.000	1090	0.11	0.002	786	-0.072	0.026	949
Doing Business Indicators: expected number of procedures it takes to start a business	-0.071	0.001	2175	0.049	0.069	1404	0.226	0.000	1880

	Number of outages in a typical month			ln El. consumption last FY (kWh)			ln Sales in last FY (USD)		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
Doing Business Indicators: expected number of days it takes to start a business	-0.038	0.075	2175	-0.461	0.000	1404	0.249	0.000	1880
Doing Business Indicators: expected cost (% of income) for starting a business	-0.041	0.057	2175	-0.464	0.000	1404	0.141	0.000	1880
Doing Business Indicators: expected capital required (% of income) to start a business	-0.023	0.275	2262	0.157	0.000	1438	0.075	0.001	1960
Doing Business Indicators: expected number of days it takes to get an electrical connection	0.118	0.000	2262	-0.146	0.000	1438	0.026	0.258	1960
Doing Business Indicators: expected number of procedures it takes to get an electrical connection	0.208	0.000	2262	-0.113	0.000	1438	-0.188	0.000	1960
Doing Business Indicators: expected cost (% of income) of getting electricity	-0.036	0.091	2262	0.064	0.015	1438	-0.11	0.000	1960
Doing Business Indicators: expected cost (USD) of getting electricity	0.099	0.000	2262	0.161	0.000	1438	-0.032	0.152	1960
Share of total population with secondary education	-0.082	0.001	1749	0.245	0.000	1224	-0.182	0.000	1502
Share of population who uses a mobile phone for banking	-0.041	0.172	1090	-0.306	0.000	786	0.473	0.000	949
Share of population with a bank account	0.165	0.000	1090	0.129	0.000	786	-0.115	0.000	949
Share of population with a mobile phone	0.052	0.084	1090	-0.104	0.004	786	0.275	0.000	949
Share of population who ever used the internet	-0.037	0.225	1090	-0.217	0.000	786	0.391	0.000	949
HDI	0.036	0.085	2262	-0.028	0.294	1438	0.013	0.569	1960
Share of population who has a partner	0.101	0.000	2167	0.018	0.511	1409	0.05	0.030	1874
Women without any formal education	0.010	0.639	2167	0.020	0.457	1409	-0.263	0.000	1874
Men without any formal education	-0.003	0.871	2167	0.029	0.280	1409	-0.265	0.000	1874
New businesses registered (number)	-0.069	0.005	1599	0.275	0.000	1074	-0.028	0.313	1342
Electric power consumption (kWh per capita)	-0.013	0.648	1268	-0.605	0.000	804	0.281	0.000	1184
Electric power transmission and	0.073	0.009	1268	-0.155	0.000	804	0.313	0.000	1184

	Number of outages in a typical month			ln El. consumption last FY (kWh)			ln Sales in last FY (USD)		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
distribution losses (% of output)									
Access to electricity. urban (% of urban population)	-0.06	0.005	2262	-0.258	0.000	1438	0.115	0.000	1960
Access to electricity. rural (% of rural population)	-0.075	0.000	2144	-0.058	0.033	1371	-0.003	0.882	1842
Access to electricity (% of population)	-0.072	0.001	2262	-0.096	0.000	1438	0.006	0.780	1960
Multidimensional poverty headcount ratio (% of total population)	0.311	0.000	469	-0.081	0.136	340	0.081	0.091	435
Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)	0.058	0.005	2262	0.145	0.000	1438	-0.139	0.000	1960
New business density (new registrations per 1.000 people ages 15-64)	-0.056	0.025	1599	-0.193	0.000	1074	-0.116	0.000	1342
Ease of doing business index (1=most business-friendly regulations)	0.071	0.001	2262	-0.126	0.000	1438	-0.097	0.000	1960
Ease of doing business score (0 = lowest performance to 100 = best performance)	-0.031	0.141	2262	0.196	0.000	1438	0.005	0.834	1960
Population living in slums (% of urban population)	0.088	0.000	2136	0.31	0.000	1348	-0.386	0.000	1834
GDP per capita (constant 2010 US$)	-0.051	0.016	2262	-0.011	0.685	1438	0.032	0.163	1960
GDP per capita. PPP (constant 2017 international $)	-0.056	0.008	2262	-0.011	0.674	1438	0.063	0.005	1960
GDP per unit of energy use (constant 2017 PPP $ per kg of oil equivalent)	0.197	0.000	1580	0.129	0.000	964	-0.113	0.000	1486
GNI per capita. PPP (constant 2017 international $)	-0.036	0.088	2262	0.004	0.888	1438	0.046	0.041	1960
Investment in energy with private participation (current US$)	-0.059	0.010	1958	0.223	0.000	1301	-0.039	0.115	1666
Firms experiencing electrical outages (% of firms)	0.059	0.005	2262	0.010	0.704	1438	-0.013	0.565	1960
Firms using banks to finance investment (% of firms)	0.094	0.000	2262	-0.16	0.000	1438	-0.172	0.000	1960
Firms using banks to finance working capital (% of firms)	0.093	0.000	2262	0.019	0.461	1438	-0.28	0.000	1960

	Number of outages in a typical month			ln El. consumption last FY (kWh)			ln Sales in last FY (USD)		
	Pearson Correlation	**Sig. (2-tailed)**	**N**	**Pearson Correlation**	**Sig. (2-tailed)**	**N**	**Pearson Correlation**	**Sig. (2-tailed)**	**N**
Power outages in firms in a typical month (number)	-0.001	0.949	2262	0.267	0.000	1438	-0.069	0.002	1960
Value lost due to electrical outages (% of sales for affected firms)	-0.036	0.088	2262	0.173	0.000	1438	0.042	0.060	1960
Fixed broadband subscriptions (per 100 people)	-0.015	0.476	2262	-0.421	0.000	1438	0.038	0.090	1960
Mobile cellular subscriptions (per 100 people)	-0.067	0.001	2262	-0.341	0.000	1438	0.193	0.000	1960
Vulnerable employment. total (% of total employment) (modeled ILO estimate)	0.088	0.000	2262	0.147	0.000	1438	-0.142	0.000	1960
FDI % of GDP	-0.061	0.004	2262	0.058	0.027	1438	-0.124	0.000	1960

13.16 Appendix: Other Discarded Modelling Attempts

While it is crucial to understand the choices made that lead to the final model, it is also relevant to understand which ideas have been discarded for which reason to get an understanding for future modelling projects about the possibilities and reasons why some options are feasible and some might not. For this reason, this section will shed some light on further ideas that could have benefitted the model, the reasons why they finally had to be discarded and, if possible, under which circumstances the approaches might still be feasible for future research.

13.16.1 Modelling Uptake on Individual and Community Level

Several logistic regression model runs were tried out for the sample on the individual level in order to assess the likelihood of a person to become a manger or owner of a microenterprise. The explanatory value of these models was, however, consistently lower than that of the household level models and the idea has therefore been discarded.

Another possible approach would have been to run a linear regression on a community level with either the share of households in a community engaged in entrepreneurial activity or the number of microenterprises in a community as dependent variable. Unfortunately, this was not possible to obtain weights for the different communities that would have been necessary to ensure representativeness.

13.16.2 Modelling for Specific Business Types

From findings in literature and own descriptive analyses, the type of enterprise seemed to be decisive in determining the level of electricity consumption and sales. It was therefore considered to be potentially beneficial to run a logistic regression not only on the uptake of an EP in general but instead on the uptake of a specific type of business. For this, the standard logistic regression model was run for the dichotomous variable if one of the household members is running a microenterprise in a specific ISIC group. Accordingly, it was considered to build each of the modes on connection, sales and consumption for the most important types of businesses as well to get more accurate results. However, there were different limitations which made It necessary to change this plan.

The resulting uptake models had decent enough Nagelkerke R^2 values around 20% but did not manage to classify the HH accurately and strongly underestimated the number of enterprises. This could possibly be explained because the types of enterprises were not well represented by the ISIC codes. Neither the HH nor the EP surveys reported the types of enterprises in a way that was found in literature. While latter was quite explicit about the type of enterprises, the ISIC codes do not fully represent the types of businesses in a detailed enough way by indicating, for instance, hairdressers, barber shops and laundry services, but only provide very generic categories. Furthermore, the independent variables might not suffice to capture the criteria of households to choose a specific type of business. The idea was therefore discarded.

The result of neglecting the business type is that it is implicitly assumed that the different types of enterprises will emerge for the same share as in the samples or at least that firms will have similar patterns for sales and electricity consumption, associated with the same predictors, across the whole population. As this is not very realistic, it is highly advised to put a major focus in future research to find ways to account for different business types by defining traits by which they can be distinguished best rather than to opt for ISIC codes. In the present research, this was attempted by conducting a cluster analysis

13.16.3 Cluster Analysis

The motivation to form clusters came from the idea that there are structural differences in consumption patterns, such as the level of consumption the likelihood to get a connection and sensitivity to outages among different business types that the ISIC codes could hint at but did not manage to represent (see Figure 50 for the difference in frequency distributions of the energy intensities (spending on electricity/sales per year) between the aggregated ISIC Rev 3.1 groups in the EP Survey).[34]

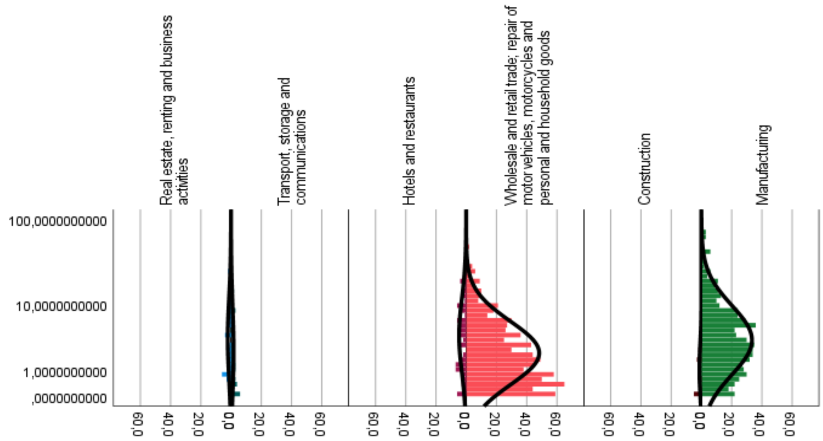

Figure 50. Frequency distributions of energy intensities (electricity cost in USD/sales in USD) for different ISIC groups in the EP Survey.

The idea was to form clusters independent of the ISIC code that would represent different styles of consumption. The ISIC codes could have been used in retrospect to see if different consumption patterns tend to be associated with certain types of industries or otherwise, a share for each cluster could have been assumed of the projected number of enterprises.

Different variables by which to cluster the EP were tested (inspired by the findings from Appendix 13.13) such as the intensity in kWh/sales, percent of losses of sales due per hour of outage and electricity cost per sales. Despite a seemingly good differentiation between clusters of different clustering attempts (see Figure 51 for an example) the regressions for sales and electricity consumption were not as strong as for the general model so the idea was discarded. A possible explanation is that the samples became too small. It is, however, still a promising idea that could be elaborated un future research.

[34] The methodology used for clustering was a combination of the next-neighbor algorithm to determine the number of clusters and the Ward method to determine cluster affiliation.

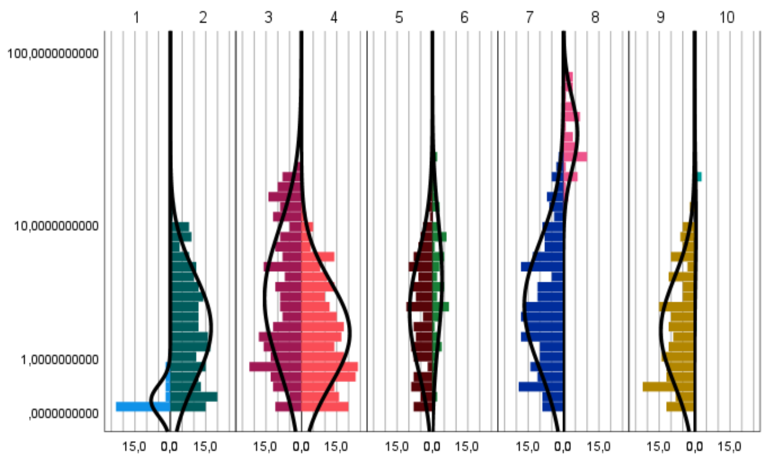

Figure 51. Frequency distributions of energy intensities (electricity cost in USD/sales in USD) for different clusterss in the EP Survey.

13.16.4 Regression on optimal consumption

Many firms stated that they faced outages, which oftentimes impaired their work, expressed in losses in sales due to outages. Since the objective of the projection is to understand what total potential demand would look like, it was attempted to calculate a variable for optimal consumption based on the share of losses in sales due to outages, thereby estimating what consumption would have been like if there had been no outages as suggested in Appendix 13.13. Testing the variable, however, did not improve the model and given the conceptual limitations encountered in the computation of the optimal consumption variable, it was decided to discard this attempt and include the number of outages as independent variable instead.

Another option would have been to only apply the model to those households which had had not faced outages. However, this was not done since it would have decreased the sample size too much and structural differences are very likely to be present in the other variables between the subset of enterprises which had faced outages and the subset of enterprises which had not faced outages.

13.16.5 Panel Regression

Analysing time-series data can give deeper insights not only in the drivers but also in the speed of uptake as a result from changing circumstances. As to this point, no longitudinal study on the uptake or consumption of microenterprises has been conducted, this would have been a valuable additional contribution. However, due to the low rate of survival in the panel studies of the World Bank Household Surveys and the absence of panel data for the Enterprise Surveys, the additional benefit was estimated to be rather low. Additionally, the longitudinal weights for the Nigerian HH Survey were still being prepared at the time of data collection.

13.16.6 Appendix: Interaction terms

It was considered to make use of interaction terms in the regression functions to represent the many interactions between predictor variables. These interaction effects are more than just the correlation between independent variables. An interaction effect takes place when a variable has different degrees of impact on the dependent variable for different values of another variable. This can lead to an effect also referred to as mediator or suppressor effect, which is a bias of coefficients so they do not represent their actual bivariate correlation with the dependent variable anymore (Lynn, 2003).

One example of such interaction effects was found in the consumption model between outages and the Doing Business scores for starting a business and getting electricity. The consumption of EPs in areas with a high score for getting electricity tends to depend less on the number of outages than the consumption of EP with a lower score. Another way to look at it would be to say that higher scores can somewhat mitigate the limiting effect outages have on consumption. At the same time, a higher electricity score only seems to encourage consumption of those businesses which face low levels of outages. For those with high levels of outages, a better score for getting electricity has no effect on the consumption. Figure 52 displays these effects by showing the correlations between the score and electricity consumption and outages and electricity consumption after grouping the cases to categories for the upper and lower 50% of values and indicating the correlations of the single groups. The same interaction effect holds between the score for starting a business and the number of outages.

However, interaction terms were excluded from the models for two reasons. First, it cannot be said for sure that the interactions which hold in the sample also hold across the entire population of SSA. Furthermore, as these interactions got quite complex (i.e. it was not just single variables which interacted but usually several variables which correlated with each other to different degrees) a very diligent selection process would have been necessary to find out how interaction terms also correlated with each other to choose the most important interactions to avoid problems such as double counting and loss of transparency. It is not always clear where these effects take place, so if not every single variable is investigated, it is easy to overlook them. The cases which were identified in this research were discovered when qualitative pre-knowledge suggested that a closer investigation of some specific variables would be wise or by chance for example when coefficients took on the reversed sign of what was expected. However, not in all cases the interaction effects reverse the sign of coefficients and the variables which interact are not necessarily the ones which are most correlated so in principle, all variables could potentially interact and would have to be tested in combination with the other variables, especially if there is not much qualitative knowledge about potential interactions. Unfortunately, this was beyond the scope of this research. Still, investigating these effects can be expected to bring very interesting insights and explain some of the contradictory findings from literature. The strategy to deal with interaction effects in this research was to ensure that the signs of coefficients were in line with the findings of the bivariate Pearson correlations, comparisons between column means and comparisons between column proportions of the independent and dependent variables (section 13.15). It was alco checked that the order of magnitude of coefficients did not change as new variables were introduced to the model.

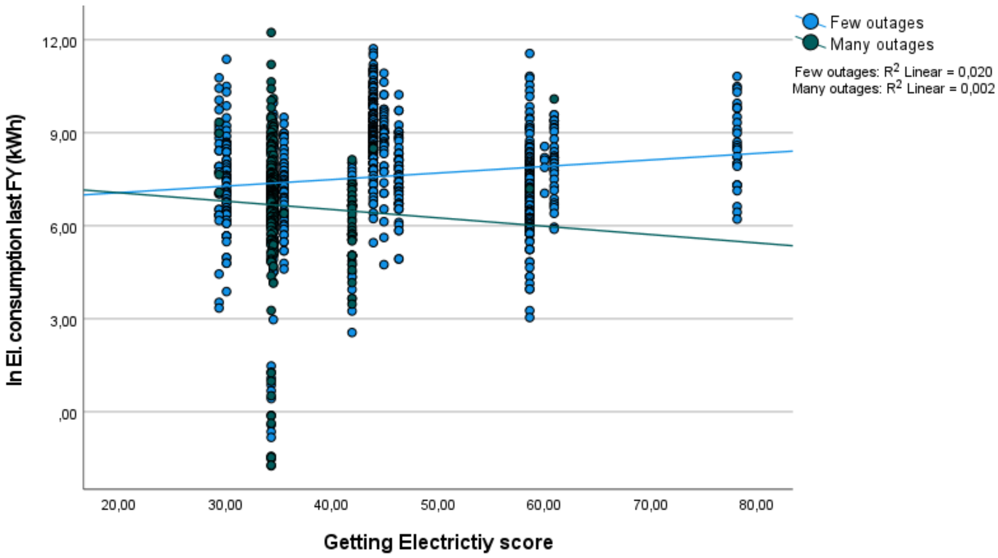

Figure 52. Correlation between score for getting electricity and electricity consumption under different levels of outages.

13.17 Appendix: Alternative Modelling Approaches

In the case that regressions fail to compute plausible results (see section 3.4), alternative modelling techniques are possible. For each of the regressions it is possible to simulate values with a Monte Carlo simulation based on a certain distribution function with values for mean, median, minimum, maximum and standard deviation based expected values from the surveys and other literature. It would furthermore be possible to exclude the sales model from the total model in case it adds more uncertainty than precision. If all values would have to be substituted by fixed shares and sales were excluded, the adjusted methodology would look like given in Figure 53.

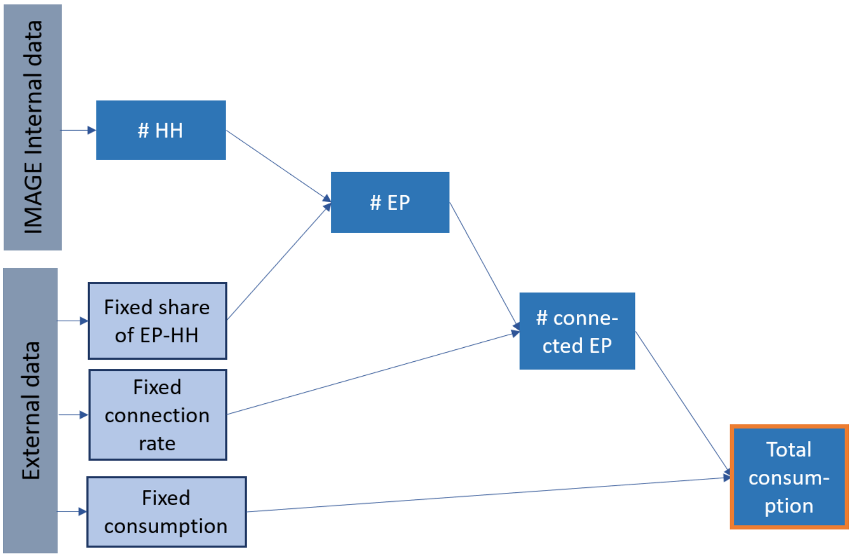

Figure 53. Alternative methodology if regression models fail.

13.18 Appendix: Acceleration of Electricity Consumption

The present section is dedicated to illustrating how the dependence of business activity on electricity accelerates electricity consumption exponentially. It is easiest to understand by using a simple example. Assume a simple hypothetical scenario with 100 separately located HH where business activity does not depend on electricity access and presence of EP amongst HH was stable at 10%, consumption per EP was 500 kWh/year and electricity access increased by 10% each year. The total consumption would be computed by #HH*share of EP(%)*share of electrified HH/%)*consumption per EP (kWh), so for this example this would give 100*0.1*0.1*500=500 kWh of total consumption for productive uses per year. For a 20% electrification rate, the total consumption would be 100*0.2*0.1*500=1000 kWh/year and so on. With increasing electrification, the consumption would increase linearly (see Figure 54). Assuming, however, that an increase in electrification by 1% would lead to an increase of EP HH of 1% (percentage point) would instead give a total consumption at a 20% electrification rate of 100*0.2*0.2*500=2000 kWh, which equals an exponential increase (see Figure 55).

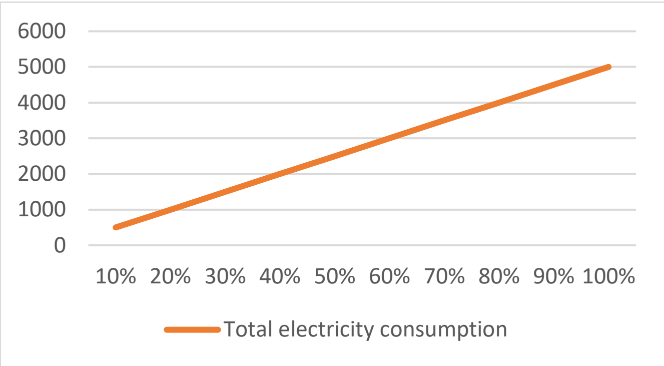

Figure 54.Total electricity consumption in hypothetical scenario of uptake did not depend on the electrification rate.

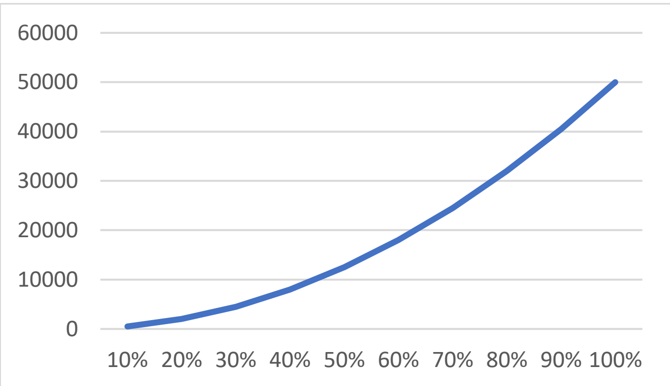

Figure 55.Total electricity consumption in hypothetical scenario if uptake deepened on the electrification rate.

This is of course, an extreme example. Sine the change in the share of one additional unit of electricity access cannot be linearly related to the increase in probability of entrepreneurial activity in the regression model, it cannot be said for sure what the actual relationship between electrification and uptake is but holding all other values at 0, the increase in probability of uptake can be calculated by increasing electrification in 10% steps. This increases the probability (and therefore the share of HH with EP) from 26.3% at 0% electrification to 36.7% at a 100% electrification rate. Assuming the approximate median observed consumption for each EP (approx. 1700 kWh/year), this gives an electricity consumption of 100 HH of 62,434 kWh at 100% electrification. At a constant share of EP of 26.3%, total consumption at universal electrification would be 46,375 kWh. If the positive relationship between electricity uptake and EP consumption had not been discovered, electricity consumption would have been underestimated by about 25% (see Figure 56).

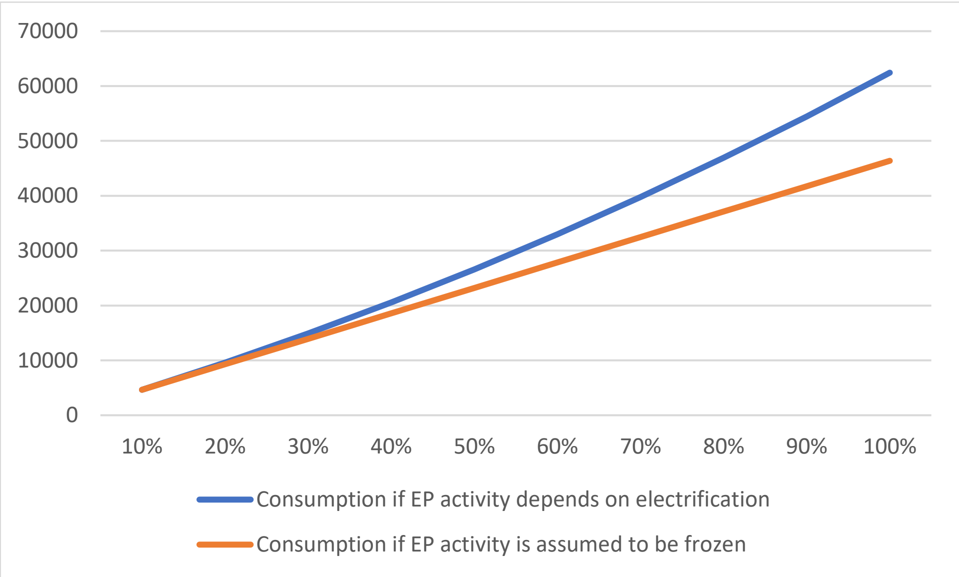

Figure 56. Total electricity consumption with increasing electrification considering dependence of EP activity on electrification and considering no dependence of EP activity on electrification.