UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Which firm characteristics determine access to finance for SMEs within the East African Community?

Emil Jonsson Joseph Fridner

Supervised by

Charles Nadeau

A thesis submitted for the degree of

Bachelor of Science in Business and Economics

11 June 2020, Spring Term

Abstract

Financial constraints among SMEs are generally more prevalent in the developing world than in the developed world, but SMEs in sub-Saharan Africa stand out as being particularly constrained. Previous studies also show causal links between access to finance and company growth and increased prosperity. By applying a logit regression model to company-level survey data of companies within the EAC, this study examines which company characteristics significantly affect access to formal external credit. The study is conducted on both the sub-regional and country level. The multiple comparisons problem is addressed by controlling the false discovery rate. On the sub-regional level, the study finds significant effects of size, managerial experience and legal status, of which size is used as a proxy for information asymmetry. Concluding policy suggestions for the EAC member nations include creating an SME database, to help close the informational gap between creditors and debtors; and providing aid to small, growing companies, to help them reach the size at which they can finance further growth through formal external credit.

For abbreviation specifications, see table A.1. For explanations regarding notation, see table A.2.

Keywords: SME, access to finance, East African Community, sub-Saharan Africa, logit, FDR, asymmetric information

Acknowledgements

We would like to extend our sincere gratitude to our supervisor, Dr. Charles Nadeau, for providing continuous guidance during the writing process.

We also want to thank Dr. Andreas Dzemski, Dr. Ann-Sofie Isaksson and Prof. Måns Söderbom for their crucial methodological advice.

We are also grateful to Prof. Arne Bigsten for providing valuable inspiration and directing us towards other people relevant for the study.

And lastly, we want to give thanks to our Lord and Savior Jesus Christ for the inspiration to create sustainable value for our fellow human beings.

Contents

List of Tables

List of Figures

1. Background

The 8th of the UN Sustainable Development Goals regards "decent work and economic growth". One of the main targets is described (UN, 2018) as being to:

> Promote development-oriented policies that support productive activities, decent job creation, entrepreneurship, creativity and innovation, and encourage the formalization and growth of micro-, small- and medium-sized enterprises, including through access to financial services.

There is no general definition of micro-, small- and medium-sized enterprises (henceforth MSMEs), as it differs a lot depending on which institution one asks. In the EU, the formal SME definition includes micro-firms and is based on number of employees and either turnover or total assets (European Union, 2003). In India, the definition is based on investments in plants, machinery and equipment, and the breaking point between the categories differs between sectors (Vandenberg et al., 2016, p. 82). When it comes to banks, the definitions used by the participants in the report by Calice et al. (2012, p. 8) differ widely[1]. Among those, the most common measures are based on loan size and company turnover, wheras number of employees, capital employed and income to bank are used less frequently. With that said, the number of employees is used extensively as the classification criterion of choice by researchers, at least when it comes to studies focusing on sub-Saharan Africa (henceforth SSA). Some of these studies also claim to focus on small- and medium-sized enterprises (henceforth SME), but choose to also refer to and discuss micro-firms or simply treat micro-firms and small firms as one category (Fjose et al., 2010, p. 5, 22, 25; Quartey et al., 2017, p. 5; Kira, 2013a, p. 54).

Regardless of the dispersion of the definition, micro-, as well as small- and medium-sized enterprises are clearly seen by the UN as important drivers of 'decent work creation and economic growth'. It should also be mentioned that SMEs make up the vast majority of businesses in the world, with the exact figure estimated to be at least 95% (Quartey et al., 2017, p. 19; Fjose et al., 2010, p. 5).

As part of the goal of helping MSMEs grow, and hence also creating economic growth and reducing poverty for the countries in which they operate (Beck & Cull, 2014, p. 584), the UN highlights access to financial services as a key success factor. As an example of policies aimed at reaching such goals, Kuwahara and Yoshino (2015, p. 3) describe how

[1] Here, the focus is solely on SMEs.

Asian countries can take inspiration from the Japanese Credit Risk Database in order to make SME data more accessible, thus lowering the degree of information asymmetry and enhancing the likelihood that banks grant credit to SMEs. More than simply a contributing factor, access to finance is instrumental for firms to be able to grow their businesses, for instance by expanding their facilities and hiring new staff (Fowowe, 2017, p. 6). It should also be highlighted that access to finance has previously been shown to have a causal relationship with entrepreneurship, company innovation and industrialization (Beck and Cull, 2014, p. 584; Fombang and Adjasi, 2018, p. 81).

This study focuses on the East African Community (henceforth EAC), which is located within sub-Saharan Africa and encompasses Rwanda, Burundi, Kenya, South Sudan, Tanzania and Uganda. With regards to the importance of SMEs for the region, the East African Community (2012) describes that

> Small and medium-sized enterprises (SMEs) represent the biggest part of all registered entities in nearly all activities in majority of the East African countries, averaging 60% in number or reaching 90% if the micro enterprises are also considered.

The reason for focusing on the EAC is that it has shown great progress in recent years. For instance, the EAC has proven to be performing well when it comes to public investment efficiency in sub-Saharan Africa (IMF, 2018, p. 76). If we narrow it down to the individual countries, there are several examples of high performance in terms of development. One example is Rwanda, which around the time of the Rwandan genocide of 1994 was the second poorest country in all of SSA, but which has since then had a more than three-fold increase in GDP per capita, as well as a decrease in poverty rate from 77% to 55.5% (in 2017) (World Bank, 2019b, p. 5). Tanzania is another country that has achieved a poverty reduction. Between 2007 and 2012, poverty rates dropped from 34.4% to 28.2% in the mainland part of the country, and in Zanzibar the poverty rate decreased from 34.9% to 30.4% between 2010 and 2015 (World Bank, 2017, p. 17).

2. Problem Discussion and Purpose

2.1 Problem discussion and analysis

Despite the importance of SMEs having access to means of financing their businesses, they are often found to be constrained in this area. Fowowe (2017, p. 6) describes that financial constraints are common among companies in developing countries in general, but are particularly prevalent among African companies. This is backed up by survey data showing that firm managers consistently rank access to finance as one of their major obstacles. This view is further expressed by Kuntchev et al. (2012, p. 13), who write that the probability of observing a 'fully credit constrained'[2] SME is higher in SSA than in many other regions. Quartey et al. (2017, p. 19-20) also stress that SMEs in SSA are in the so-called "missing middle", in the sense that they are too large to get micro-finance loans and too small to possess the necessary means of obtaining financing from banks. Further, they also highlight research establishing that "most SME loan applications in Africa are not granted", and that these companies instead rely on retained earnings and informal lending.

If we narrow it down to the EAC, the situation is consistent with that of the rest of SSA. Kira (2013a, p. 63) concludes that within the EAC,

> [...] most of the firms experiencing financing obstacles tended to possess SMEs' features i.e. small and medium, young, sole proprietorship and partnership [...].

Two of the main reasons for this scarcity are a high degree of information asymmetry and a lack of collateral (Kira, 2013a, p. 50, 52; Mugwika, 2013, p. 36; Quartey et al., 2017, p. 19-21, 26). Information asymmetry is also an obstacle when it comes to getting listed on a stock exchange. As such, in terms of formal financing, SMEs are left to rely on banks (Kira, 2013a, p. 52-53). One problem is that banks also like to have information about their borrowers. For example, when banks in SSA were surveyed about which factors associated with SMEs that they considered to be essential obstacles to issuing loans, 100% of the respondents in Kenya and Uganda, and 75% in Tanzania, chose a lack of information (Calice et al., 2012, p. 26-7).

[2]The definition of being 'fully credit constrained' is discussed further in section 5.

AfDB (2012, p. 5) finds that the key factors of importance for firms' prospects of accessing formal external credit include "the business plan, projection of monthly cash flows, existence of audited book accounts and a track record of banking transactions". All of these can reasonably be described as enhancing potential creditors' information about a given company. In Rwanda, it was also found that having a bank account and regular book of account were well established among the surveyed SMEs, whilst 70% of them did not have a business plan and 72% lacked cash flow projections (AfDB, 2012, p. 6).

According to Chavez (2017, p. 1, 3, 12), a way of mitigating the problem with asymmetric information is to implement a "credit information system, where credit reporting service providers (CRSP) collect and distribute data on the repayment histories of borrowers [...]". However, SSA possesses the least developed credit reporting system according to data from the region.

Another reason for companies in the developing world being credit constrained, is that there is a lack of such institutions as private equity- and venture capital firms that focus on SMEs. Even when they do exist, SMEs are often unable to meet the expectations of the investing party, e.g. in terms of preparedness for the due diligence process (Divakaran et al., 2014, p. 6, 12).

2.1.1 SME definition

Most of the research on the topic only explicitly discusses SMEs, and to ensure that this is applicable and comparable, micro-firms are excluded from the study. To be able to do this, there is a need to define small- and medium-sized enterprises. We define the term based on the aforementioned criterion of number of employees, as we have access to that information from our data source, the World Bank Enterprise Surveys (henceforth WBES). Using the definition by the European Union, for instance, may risk introducing bias. This is because we can reasonably expect that a European SME differs from an East African SME. To make our definition locally applicable, we utilize the definitions used in the six EAC member nations. By taking the median of the upper limits for micro firms and the lower limits for large firms, we derive our cut-off points (Ministry of Trade, Industry and Cooperatives, 2015, p. 25; World Bank, 2014) (see table A.9).

2.2 Purpose and research question

As previously stated, accessible financing has been shown to increase growth, job creation and prosperity, and is emphasized in the UN's 8th Development Goal. Hence, the proven obstacles SMEs in the region face when attempting to obtain external financing are an important issue to address and solve. Thus, the purpose of this thesis is to further examine which firms characteristics significantly impact access to finance. In doing so, the goal is to shed light on what can be done by policy makers to address these characteristics, in order to ultimately solve the SME financing problem and reap the associated societal benefits. As a result, the research question is specified as:

> *Which firm characteristics determine access to finance for SMEs within the EAC?*

It should be noted that this implies that only the directions of the determinants' effects are of interest. As such, their magnitudes are not analyzed.

2.2.1 Main contributions

The study is expected to contribute to the research field by

- adding to the growing literature about SME financing in SSA in general and the EAC in particular;
- providing an updated examination of SME financing in the EAC using the WBES, due to having access to more recent data sets and including the new Community member South Sudan in the analysis;
- utilizing an objective measure of access to finance[3];
- including variables which previous studies of the EAC omit[4]; and
- controlling the false discovery rate using the method by Benjamini and Hochberg (1995, p. 291).

[3]Previous studies by e.g. Kira (2013a) sometimes use a subjective measure of access to finance based on company perceptions.

[4]This includes variables such as $mang\_fem$, for_owner, *exporter* and *audit*, which are discussed further in section 5.

2.2.2 Delimitations

As part of the study's delimitations, we

- exclude micro- and large enterprises as well as agricultural firms;
- use cross sectional data instead of panel data, due to data limitations;
- focus specifically on firm characteristics when identifying determinants of access to finance, to stay consistent with the purpose of the study; and
- focus only on debt financing, as financial markets for equity financing are not very extensive in the developing world.

3. Literature review

On the topic of access to finance, the available literature examines both African and non-African countries. Studies of sub-Saharan Africa often focus on a few countries, an economic community or some other region within SSA. SMEs being credit constrained is usually treated as an axiom, for which the main task is finding determining factors. Particular emphasis is also sometimes put on how access to finance affects gender equality, innovation and other themes alike. The set of variables with a significant positive or negative effect usually differs from country to country, and from region to region within SSA.

When it comes to which firm characteristics affect access to finance, Quartey et al. (2017, p. 23) describe that within ECOWAS, the possibility of obtaining financing is connected to factors such as firm size, firm age and experience of the top manager. They also find that the set of significant explanatory variables differs between countries within ECOWAS. In Ghana, for example, firm size and formality are identified as top drivers of access to finance, whereas in Senegal, the set of significant variables includes ownership and firm performance. Kira (2013a) conducts a study of the determinants of access to finance within the EAC using WBES data, although with numerous methodological differences[5] in comparison to this study. He finds that access to finance is significantly affected by variables such as company size, company age, legal status and economic sector (Kira, 2013a, p. 63-4). The study conducted by Kuntchev et al. (2012, p. 17-18) also supports the significance of firm size as an determinant. Other studies find significant effects of variables such as GNP per capita, population density, manager experience and whether the firm is an exporter or owned by foreigners (Aterido et al., 2013, p. 105; Beck and Cull, 2014, p. 599; Fombang and Adjasi, 2018, p. 78). The proxy for access to finance varies between studies. A common measure, used by Aterido et al. (2013) and Beck and Cull (2014), is a simple binary variable for whether the company has a loan[6]. Other studies use both objective and subjective measures. Quartey et al. (2017, p. 21-22), for instance, use an objective measure, namely the proportion of working capital financed with internal funds, to conduct their main analysis. For the purpose of robustness checks, they then use 3 subjective measures based on company perception regarding how severe of an obstacle access to finance is. Another objective measure developed by Kuntchev

[5]This includes using the ordered probit model, a subjective response variable, a different regression specification and a different sample.

[6]The exact definition differs slightly. Aterido et al. (2013) also include access to overdraft, for example.

et al. (2012, p. 9-12), also used by Fowowe (2017, p. 10), is a measurement scale of the degree of credit constraint, based on the WBES. This scale consists of 4 categories, with certain criteria connected to each category. A simplified description of the division scheme is given in figure A.1. This objective measure is then used to create the ordinal response variable *credit constraint status* (CCS)[7].

According to the available literature, increased SME financing has several positive consequences, both for the company and for the economy. For example, Fowowe (2017, p. 16) finds that if a company participates in the financial market, the company's growth is enhanced. Beck and Cull (2014, p. 584, 599) further discuss how financial deepening has a positive effect on reducing poverty, and that SME financing plays an important role in the matter. It is also brought to attention that accessible external financing is positively correlated with the number of start-ups, and hence the degree of innovation. When examining this connection between access to finance and innovation, Fombang and Adjasi (2018, p. 84) conclude, based on their sample[8], that "access to finance (through trade credit, asset finance and overdraft) enhances innovation in the respective country".

With regards to formulating the methodology of this study, the literature used consists mostly of the oft-cited textbook by Wooldridge (2015) and of publicly available lecture notes by Söderbom (2009a), Professor of Economics and Head of the Department of Economics at the Gothenburg School of Business, Economics and Law. Also, the methodology by Kuntchev et al. (2012) is referred to for constructing the response variable. In addition, the oft-cited study by Benjamini and Hochberg (1995) is used for the purpose of robustness tests, and studies by Streiner (2015) and Mundfrom et al. (2006) are used more generally to formulate the problem of multiple comparisons.

Just like a great deal of the available literature on the subject, this study is based on WBES data which, as stated by Kuntchev et al. (2012, p. 5), is used by hundreds of academic papers. The major differences between this study and other studies concern the response variable and the robustness tests. To our knowledge, we are alone in using a binary response variable based on that by Kuntchev et al. (2012, p. 9-12), and in controlling the false discovery rate using the method by Benjamini and Hochberg (1995, p. 291) (see section 6.3). An additional difference is that some studies also include subjective response variables in their analyses, variables that are based on the respondents' perceptions. This study however, does not include such subjective measures.

[7] This methodology is expanded upon in section 5.
[8] Consisting of Kenya, South Africa, Cameroon, Nigeria and Morocco ($n = 5304$).

4. Theoretical Framework

An important theory that applies to SME financing is that of asymmetric information and the market failures that follow from it, famously formulated by Akerlof (1970). This is discussed in several studies, such as Chavez (2017, p. 1-2, 12), Quartey et al. (2017, p. 22, 26) and Kira (2013a, p. 52). The theory itself concerns an economic transaction in which one party has more information than the other party has. Akerlof (1970) exemplifies this with the sale of a used car, in which the seller possesses more information about the car than the buyer does, making the buyer set his maximum price to that of a car of average quality, even when the car is in fact of top quality. This occurs because the seller has no way of credibly conveying the car's quality to the buyer. In the same way, SMEs suffer from limited means of conveying their credit worthiness to financial institutions (Calice et al., 2012, p. 26-7). As such, these institutions are expected to charge higher risk adjusted interest rates, compared to the rates charged from the SMEs' more sizable peers, leading to a shortage of credit for these SMEs. It should be mentioned that in this study, moral hazard is not touched upon, so the terms 'information asymmetry' and 'adverse selection' are used more or less synonymously.

In previous research, various different variables are used as proxies for information asymmetry. Examples include the share of intangible assets, used by Fidrmuc et al. (2015, p. 16), and a proxy variable derived from "analysts' forecasts and stock price reactions to earnings announcements", used by Thomas (2002, p. 2). Kalash and Sabsabi (2018, p. 443) use firm size, defined as the natural logarithm of total assets, as a proxy for asymmetric information. Gertler (1988, p. 23-4) discusses how information asymmetry is tied to company age, stating that in the presence of such asymmetries, young companies may find it more difficult to obtain financing from lenders. In addition, Chavez (2017, p. 12, 19) uses whether the firm has been externally audited as a proxy for firm transparency, and shows that firms with a lower degree of transparency have more to gain from accessible credit information in the market.

5. Data

5.1 The Sample

Ideally, answering the previously stated research question would include examining financial statements. However, due to the nature of SMEs in the developing world, those are hard to come by and in many cases nonexistent (Sacerdoti, 2005, p. 8). Instead, this study is based on the WBES, an ongoing data-collection project conducted by the World Bank, as part of its "goal of building a climate for investment, job creation and sustainable growth" (Kuntchev et al., 2012, p. 4). The data is collected from a wide selection of countries, currently encompasses $> 130,000$ companies and is both objective and subjective in nature. It is collected using a standardized methodology, on a 3-4 year basis in each region. The standardization of the data makes it useful for comparisons between different time-periods and countries. Further, the data obtained is from non-agricultural firms[9] within the formal and private economy, which can broadly be divided into service and manufacturing. Finally, it is worth noting that in the data collection process, the population of companies is divided into sub-populations based on the number of employees, and the sample is then collected from each sub-population (Kuntchev et al., 2012, p. 4-7). This is to help ensure that companies of different sizes are all represented in the sample[10].

5.1.1 Our sample from the WBES

Our sample consists of firm-level data from Rwanda (2011)[11], Uganda (2013), Burundi (2014), Tanzania (2013), Kenya (2018) and South-Sudan (2014) (East African Community, 2020a). The sample is tabulated further in table A.3. In cases where a country has been surveyed multiple times at different points in time, only the latest data set is used, to avoid observing the same company multiple times[12].

[9] Agricultural firms make up a major sector in the EAC (East African Community, 2018), and if there was available WBES data on these firms, then it would have been included.

[10] This is known as stratified sampling.

[11] Parentheses indicate when the data was collected in the different countries.

[12] Within the WBES, It is not possible to reliably identify a specific company in two different data sets. Hence, using multiple datasets from one country would bring the risk of introducing bias.

5.2 Cleaning the data

The first and most obvious measure taken to clean the data is to exclude all observations for which data is missing for ≥ 1 of the explanatory variables. Observations are also excluded if data is missing for ≥ 1 of the intermediary variables used f.e. to create the response variable. Further, to stay consistent with the purpose of the study, all companies with outstanding publicly traded stocks are excluded. The study also excludes all companies using informal credit, so as to isolate the effect on access to formal external financing. In order to only include small- and medium-sized enterprises in the sample, all companies not defined as such are excluded, according to the approach outlined in section 2.1.1. The cut-off point obtained is $5 \leq$ employees ≤ 100 for SMEs in the EAC (see Table A.9)[13].

5.3 Variables

5.3.1 Access to Finance

The choice of proxy for access to finance is adapted from the measurement created by Kuntchev et al. (2012), later used by Fowowe (2017). It is purpose-built based on the data available in the WBES, and is created by dividing the companies into four groups, labeled FCC (Fully credit constrained), PCC (Partially Credit Constrained), MCC (Maybe Credit Constrained) and NCC (Not Credit Constrained). This is done based on answers provided to the questions in the WBES. For a simplified overview of the division scheme, see figure A.1.

Kuntchev et al. (2012) then create the ordinal variable CCS (Credit Constraint Status) which takes on the value 1 if the company is in the NCC group, 2 if in MCC, 3 if in PCC and 4 if in FCC. However, using this measure would require an ordinal regression model, such as ordinal logit (Söderbom, 2009b, p. 1). This would complicate the study. To keep it relatively simple, the response variable is adapted by way of collapsing the four groups into two. This way, simpler regression models are applicable. The binary variable

[13]The figures for Kenya and Uganda are from Ministry of Trade, Industry and Cooperatives (2015, p. 25), since World Bank (2014) does not contain all necessary data for those countries.

CCS_bin is constructed as follows:

$$CCS\_bin = \begin{cases} 1, & \Leftrightarrow \text{MCC} = 1 \text{ or NCC} = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

As such, following the logic of Fowowe (2017, p. 10)[14], a higher value of CCS_bin entails a higher degree of access to finance. In words, this means that CCS_bin takes on the value 1 if the company has been granted a loan, or if it has refrained from applying for a loan due to already having sufficient funding; and takes on the value 0 otherwise. Descriptive statistics for the sample, including CCS_bin, are shown in table A.3.

5.3.2 Explanatory variables

The explanatory variables are defined in table A.7. The chosen explanatory variables are ones which have been included in previous research, many of which have been found to have a significant effect on access to finance, as can also be seen in table A.7. Further, we include country fixed effects and $lgCity$ as control variables. Descriptive statistics are provided in table A.4. Worth noting is that firms being run as partnerships are used as the benchmark group for the $LS\_private$ and LS_sole dummies. Company age and size, along with $audit$, are included as proxies for information asymmetry. As mentioned in section 4, size has previously been used as such, but has been based on total assets. Since that information is not available in the WBES, size is instead based on the number of employees. The inclusion of $audit$ is due to its connection with company transparency, also stated in section 4, with the expected effect that having audited financial statements is beneficial for SMEs in terms of access to finance. To model expected nonlinear effects of age, size and manager experience, the natural logarithms of these are used as regressors.

[14]This is done for the sake of interpretational ease, so that a positive coefficient entails a positive effect on access to finance. This is the opposite of the logic of Kuntchev et al. (2012, p. 14).

6. Method

In the aim of providing more nuanced results, with greater room for analysis, the regression analysis is conducted in two steps. Initially, the entire sample is analyzed at once to identify sub-region-wide determinants of access to finance. Following this, each country is analyzed individually to identify country-wide determinants.

Having constructed the binary variable CCS_bin, the latent variable framework described by Wooldridge (2015, p. 526-7) and Söderbom (2009a, p. 8-12) is used. In other words, we assume that CCS_bin is in fact driven by an underlying continuous variable y^*, which is not observed; only the binary outcome CCS_bin is observed. In this framework, y^* is referred to as the latent variable. For the purpose of this study, it is useful to define y^* as

$$y^* = \text{Available Internal \& External Financing} - \text{Capital Need} = \text{Net Financing}, \quad (6.1)$$

where $y^* \geq 0$ means the company has enough funding to run its business, and $y^* < 0$ means it is constrained in terms of financing[15]. For future reference, we also define that $\boldsymbol{x\beta} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$[16].

Following the methodology of Söderbom (2009a, p. 10-1), we now assume that y^* is determined by

$$y^* = \boldsymbol{x\beta} + \varepsilon, \quad (6.2)$$

where ε is an error term, assumed uncorrelated with $\boldsymbol{x}$, and that the effect of y^* on the observed variable CCS_bin is defined[17] as

$$CCS\_bin = \begin{cases} 1, & \Leftrightarrow y^* \geq 0 \\ 0, & \Leftrightarrow y^* < 0. \end{cases} \quad (6.3)$$

To be clear, we would hence expect that if there is positive net financing, then $CCS\_bin = 1$, meaning the company is not credit constrained, and vice versa. From here, it can be derived that the appropriate model to use for identifying determinants

[15] If the sign on net financing is inverted, it can also be interpreted as a financing gap.

[16] Note, this can also be described as $\boldsymbol{x}$ being a $N \times K$ matrix of explanatory variables, and $\boldsymbol{\beta}$ being a $K \times 1$ vector of parameters (Söderbom, 2009a, p. 4).

[17] The inequality signs have been slightly altered from the original source to conform with the logic of how y^* is defined here, but the effect on the reasoning is minuscule.

of $\Pr(CCS\_bin = 1)$, is the logit regression model. For more details, see the discussion surrounding equation B.1 in appendix B.

As a result, it could be said that the commonly used OLS regression model

$$y = \boldsymbol{x\beta} + \varepsilon \tag{6.4}$$

is adapted to

$$\Pr(CCS\_bin = 1) = G(\boldsymbol{x\beta}), \tag{6.5}$$

where instead of the function G[18], we insert the logistic function $\Lambda(\boldsymbol{x\beta}) = \frac{exp(\boldsymbol{x\beta})}{1+\exp(\boldsymbol{x\beta})}$, such that

$$\Pr(CCS\_bin = 1) = \Lambda(\boldsymbol{x\beta}) = \frac{exp(\boldsymbol{x\beta})}{1 + \exp(\boldsymbol{x\beta})}. \tag{6.6}$$

This can then be rewritten[19] to

$$\log\left(\frac{P}{1-P}\right) = \boldsymbol{x\beta}, \tag{6.7}$$

which is the specification most useful to think of when looking at logit regression results. Since the coefficients are now defined in terms of log(odds)[20], they are naturally more difficult to interpret, in comparison with OLS coefficients. For the sake of slightly greater ease of interpretation, the regression tables report average marginal effects rather than regression coefficients[21].

The reader should note that binary response variables could also be modeled with OLS, but that this approach has several drawbacks. It could for example result in estimates where $\widehat{\Pr} \notin (0, 1)$(Wooldridge, 2015, p. 226), wheras logit ensures that $\widehat{\Pr} \in (0, 1)$. Another issue is that it does not make sense to have a probability be linearly dependent on a continuous variable for all possible values (Söderbom, 2009a, p. 6). This is also rectified by the nonlinear nature of logit, as can be inferred from equation 6.7.

With reference to the beginning of this section, the model in equation 6.6 is run using the explanatory variables in table A.7, first at the sub-regional level, and then at the country-level.

[18] While logit regression is defined by $G(\boldsymbol{x\beta}) = \Lambda(\boldsymbol{x\beta}) = \frac{\exp(\boldsymbol{x\beta})}{1+\exp(\boldsymbol{x\beta})}$, other regression models use different G functions. Probit, for example, is defined by $G(\boldsymbol{x\beta}) = \Phi(\boldsymbol{x\beta})$ where Φ denotes the standard normal CDF.

[19] For details on how this is rewritten, see equation B.2

[20] For more details, see equation B.3.

[21] For more details on marginal effects, see equations B.4 and B.5

For the sake of completeness, probit regression results are also reported.

6.1 Omitted variable bias

In causal inference, there is a tangible risk of drawing invalid conclusions due to having misspecified the model by excluding ≥ 1 important covariates. In the case of a general OLS regression, it can be shown (Wooldridge, 2015, p. 102-3) that

$$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j, \tag{6.8}$$

where $\hat{\beta}_j,\ j = 0, 1, \ldots, k$ is the OLS estimator from a regression using the full set of explanatory variables, $\tilde{\beta}_j,\ j = 1, 0, \ldots, k-1$ is the OLS estimator from the regression that excludes x_k, and $\tilde{\delta}_j$ is the slope coefficient from regressing x_{ik} on $x_{i1}, x_{i2}, \ldots, x_{ik}$. As such, the omitted variable bias term $\hat{\beta}_k \tilde{\delta}_j$ depends on the real effect of x_k on y, as well as the covariance between x_k and each x_j; if either $\hat{\beta}_k = 0$ or $\tilde{\delta}_j = 0$, then the coefficient bias is also 0. This works in a similar way in logit regression, with only a few differences, at least when it comes to determining conditions for $\hat{\beta}_j$ switching sign[22].

6.2 Reverse causality

Another point worth bringing up, that is relevant for interpreting the regression results, is reverse causality. This can be an issue for example in epidemiological studies, in which it may be the case that an ailment in fact *causes* changes in the hypothesized risk factor, as opposed to the other way around. For instance, Werner et al. (2007, p. 1376-7) find that the risk of schizophrenia is higher among people of low social status, when it may instead be that schizophrenics are more likely to end up in a lower social status. In the same way, it could be the case that if a significant effect of *lsize* is found, then this is not because *lsize* is a driver of access to finance, but that it is *driven by* access to finance. In other words, it would not be unreasonable to simply expect that a well-funded business will grow to a greater size than a poorly funded business will, and this would then potentially complicate the causal inference. One way of addressing the problem would be to instead use lagged values of *lsize*, but limitations in the availability of data render this method an impossibility.

[22] For more details on omitted variable bias in logit regressions, see the discussion surrounding equation B.6.

6.3 Robustness tests

To detect possible multicollinearity, two common methods are used, those being the use of a correlation matrix and VIF_j[23]. The correlation matrix uses the common cut-off value of $|0.8|$ (Kennedy, 2008, p.196). The VIF_j uses a cut-off value of 10, and is computed as $\left(1 - R_j^2\right)^{-1}$ (Kennedy, 2008, p.199) where R_j^2 denotes the R^2 obtained from regressing x_j on the other regressors. $\text{VIF}_j > 10$ would then be interpreted as an indication of possibly harmful multicollinearity.

When carrying out a regression analysis and testing the significance of multiple coefficients, rather than a single coefficient, there is a risk that some of these coefficients will show a significant effect purely by chance, even though the real effect is close to 0. This is known as the multiple comparisons problem and unless corrections are made, this risk increases with the number of hypotheses tested (Streiner, 2015, p. 721-2). For example, if a researcher tests 10 different hypotheses, each at a significance level of 5%, he or she will implicitly accept a $1 - (0.95)^{10} \approx 40\%$ risk of falsely rejecting ≥ 1 null hypothesis. Some methods that try to rectify this include the conservative Bonferroni[24] correction (Mundfrom et al., 2006, p. 6) and the more powerful Bonferroni-Holm correction (Streiner, 2015, p. 724). However, these can become very conservative when testing many hypotheses, and do not take into account the *number* of erroneous rejections (Benjamini & Hochberg, 1995, p. 290). This study instead applies the correction developed by Benjamini and Hochberg (1995, p. 291), which controls the false discovery rate (FDR), defined as

$$\text{FDR} = E\left[\frac{\text{Number of incorrectly rejected null hypotheses}}{\text{Number of rejected null hypotheses}}\right]. \tag{6.9}$$

The FDR is controlled by doing as follows. For null hypotheses $H_1 \dots H_m$, the corresponding p-values $p_1 \dots p_m$ are listed in ascending order and denoted $p_{(1)} \dots p_{(m)}$. Then, given a significance level q, u is defined as the largest i for which $p_{(i)} \leq \frac{i}{m}q$, after which all $H_1 \dots H_u$ are rejected (Benjamini & Hochberg, 1995, p. 293). This ensures that $FDR \leq q$. Since studies of a more exploratory nature can use a slightly more liberal level of q than would otherwise be used for the level of α for individual significance tests (Diz et al., 2011, p. 8), this study uses $q = 0.1$. Since the regression model is applied both on the sub-regional level and the country level, and we are interested in finding significant covariates on either level, the FDR correction is run once on all hypothesis test outcomes.

[23] To keep the notation consistent, the subscript j is changed from the i used in the original source

[24] The Bonferroni correction simply tests m hypotheses each at a significance level of α/m.

It is also important to note that subsequent conclusions are based mainly on the results that are not excluded upon controlling the FDR.

7. Results

After applying the previously specified regression model on the entire sample, the results in table 7.1 are obtained. For the EAC, Kenya, Tanzania, and Uganda regressions, the χ^2 tests of overall significance yield p-values < 0.01, meaning that for each of these regressions, we can determine that at least one $\beta_j \neq 0$.

On the sub-regional level, it is estimated that on the 5% significance level, company age has a positive effect on access to finance, while on the 1% significance level, a positive effect of size is estimated. A χ^2-test of the joint significance of legal status yields $\chi^2_{(2)} = 8.84$ with $p = 0.012 < 0.05$, meaning that the effect of legal status is significantly different from 0. The model estimates that being either a limited liability company or sole proprietorship, as opposed to a partnership, has a significant positive effect on access to finance. Worth noting is that the point estimate on $LS\_private$ is greater than that on LS_sole. Managerial experience is estimated to have a negative effect on access to finance on the 5% level. No significant effects of $mang\_fem$, $exporter$, for_owner and $audit$ are estimated, and hence it cannot be determined that these differ from 0.

When looking at the country-level, it should be noted that the χ^2-tests of overall significance do not reject for Burundi, Rwanda and South Sudan. For the other countries however, the χ^2-tests do reject, making it valid to interpret the individual point estimates. In the case of Kenya, estimated significant effects include $lage$ on the 1% level, LS[25] on the 5% level and and $lmang\_exp$ on the 10% level. Within Tanzania, $lsize$ is found significant at the 1% level, while $lmang\_exp$ and $audit$ are found significant at the 10% level, but no effect of LS[26] is found. Within Uganda, $lsize$ is significant at the 10% level, and $lmang\_exp$ and for_owner are found significant on the 5% level. However, no effect of LS[27] is found within Uganda.

Probit results of these regression specifications are provided in table A.8.

[25] The χ^2-test of joint significance yields $\chi^2_{(2)} = 6.90$ with $p = 0.032 < 0.05$.
[26] The χ^2-test of joint significance yields $\chi^2_{(2)} = 4.47$ with $p = 0.107 > 0.05$.
[27] The χ^2 test of joint significance yields $\chi^2_{(2)} = 2.90$ with $p = 0.234 > 0.05$.

Dependent Variable							
CCS_bin	EAC	Burundi	Kenya	Rwanda	South Sudan	Tanzania	Uganda
lage	0.036^{**}	-0.077	0.106^{***}_{τ}	-0.010	0.003	0.089^{*}	-0.032
	(0.016)	(0.066)	(0.026)	(0.053)	(0.031)	(0.053)	(0.042)
lsize	0.071^{***}_{τ}	0.163^{**}	0.030	-0.039	0.081^{*}	0.186^{***}_{τ}	0.068^{*}
	(0.015)	(0.065)	(0.024)	(0.047)	(0.042)	(0.035)	(0.037)
lmang_exp	-0.042^{**}_{τ}	-0.055	-0.053^{*}	0.013	-0.007	-0.082^{*}	-0.103^{**}
	(0.017)	(0.068)	(0.029)	(0.062)	(0.031)	(0.049)	(0.044)
mang_fem	0.026	-0.028	0.022	-0.050	0.028	0.100	0.081
	(0.031)	(0.129)	(0.052)	(0.098)	(0.070)	(0.090)	(0.066)
exporter	0.078	-0.175	0.064	-0.128	0.203	-0.083	0.162
	(0.049)	(0.169)	(0.062)	(0.219)	(0.162)	(0.123)	(0.125)
for_owner	-0.014	0.034	-0.085	0.163	-0.048	0.161	0.182^{**}
	(0.031)	(0.158)	(0.060)	(0.114)	(0.047)	(0.141)	(0.090)
audit	0.020	0.025	0.079	0.196^{**}	-0.021	-0.115^{*}	-0.010
	(0.025)	(0.096)	(0.048)	(0.083)	(0.050)	(0.066)	(0.055)
LS_private	0.086^{***}_{τ}	-0.085	0.140^{***}	0.116	-0.090	0.168^{*}	0.222
	(0.033)	(0.219)	(0.052)	(0.111)	(0.070)	(0.102)	(0.170)
LS_sole	0.062^{**}_{τ}	-0.149	0.045	0.257^{***}	0.026	0.168^{**}	0.090
	(0.026)	(0.198)	(0.045)	(0.083)	(0.048)	(0.083)	(0.064)
lgCity	-0.001	0.031	-0.022	-0.137	-0.065	0.136^{*}	0.010
	(0.024)	(0.115)	(0.040)	(0.306)	(0.048)	(0.082)	(0.053)
Country FE	Yes	No	No	No	No	No	No
N	2050	103	633	156	533	253	372
pseudo R^2	0.0613	0.0828	0.0508	0.0657	0.0159	0.1280	0.0490
χ^2	160.90^{***}	12.98	36.10^{***}	13.51	10.17	25.33^{***}	25.13^{***}

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Reported values are APEs, as opposed to coefficients

τ denotes that the variable is found to be significant after controlling the FDR

Table 7.1: Sub-regional and country-level logit regression results

7.1 Robustness tests

The correlation matrix is tabulated in table A.6. The VIF results are shown in table A.5. No correlation coefficients are found to exceed |0.8| and no VIF values are found to exceed 10. As such, no signs of harmful multicollinearity are detected.

The results from running the FDR control by Benjamini and Hochberg (1995) are also shown in table 7.1, where each significant variable is denoted by a subscript τ. After controlling the FDR, all significant sub-regional covariates, apart from *lage*, remain significant. On the country level, significant effects are now estimated of *lage* in Kenya, and of *lsize* in Tanzania.

8. Analysis

8.1 Sub-regional level

Due to how CCS_bin is defined, having estimated significant positive effects of size and legal status implies that these firm characteristics increase the likelihood that a given company either has a bank loan or has refrained from applying for a loan due to already having enough funding. The significant negative effect of manager experience implies that if the company's manager has worked for a long time in the industry, this will increase the likelihood that the company gets its loan application denied, or refrains from applying due to the associated terms and conditions. Having excluded *lage* after controlling the FDR implies that we cannot interpret *lage* as a significant sub-regional covariate without also expecting that $> 10\%$ of rejections are erroneous. Since no significant effects of $mang\_fem$, $exporter$, for_owner and $audit$ are found on the sub-regional level, no conclusions can be drawn based on either of these variables.

8.1.1 Asymmetric information

lsize and *lage*, in addition to *audit*, act as proxies for information asymmetry, as explained in section 5.3.2.

Finding a significant positive effect of size on access to finance largely confirms the findings of other studies looking at the world in general (Chavez, 2017, p. 5, 12), the developing world (Kuntchev et al., 2012, p. 2-3), SSA (Quartey et al., 2017, p. 23) and the EAC in particular (Kira, 2013a, p. 60). This is also consistent with the idea of SMEs being associated with a great deal of information asymmetry, as well as with the expectation that as they grow, they have an easier time getting access to finance. This is explained by bigger companies being more likely to possess the collateral and track record necessary to obtain external financing (Kira, 2013a, p. 50; Quartey et al., 2017, p. 23). However, it is important to also keep the risk of reverse causality in mind when making these interpretations as it could be the case that *lsize* is also driven by access to finance (see section 6.2). This is difficult to test though and is beyond the scope of this study.

With regards to firm age, after having controlled the FDR, this study's findings conform somewhat to those of Kuntchev et al. (2012, p. 2-3)[28], who could not identify such a

[28] Based on their WBES sample of 113 different developing countries.

relationship after controlling for size. However, when looking at ECOWAS, Quartey et al. (2017, p.23) write that,

> We find that firm's age significantly increases firm's access to finance from external sources and this buttresses the fact that older firms are less likely to have difficulties in accessing finance compared to newer firms.

If we narrow it down further to our area of interest, the EAC, Kira (2013a, p. 60) concludes that firms older than 10 years are less likely to face financing constraints than younger firms are. He also elaborates that there is no proven significant difference in access to finance between firm 1-4 years old and firms aged 5-9 years. It could be that the reason for our findings diverging from those of Kira (2013a), despite looking at a similar sample, is that we use a different regression specification, or that Kira (2013a) does not control for the multiple comparisons problem. Either way, *lage* should be interpreted with caution due to it being excluded upon controlling the FDR, but it is still relevant to further examine the effects of age in the EAC, since other studies have made contradictory findings.

The H_0 of *audit* not rejecting even on the 10% level could possibly result from *audit* being relatively strongly correlated[29] with *lsize*, as seen in table A.6. As such, the reason could be that information asymmetry in fact has a very significant effect on access to finance, but that this effect is already captured in *lsize*. In fact, depending on the regulation in each country, financial auditing may be required only for certain companies. For example, Kenyan companies are exempt from auditing requirements provided they do not exceed certain thresholds regarding turnover and total asset value (Republic of Kenya, 2015, p. 777). Hence, since larger companies are already more likely to have an auditor, it is likely that observing *audit* separately does not add enough information for it to be considered a significant covariate.

In summary, since the effect of *lsize* is found to be highly significant, even after controlling the FDR, there is some evidence that access to finance for SMEs in the EAC can be improved by reducing the degree of information asymmetry. However, there is still some ambiguity present based on the outcome of the *lage* and *audit* variables.

[29]Do keep in mind however, that this correlation is not strong enough to exceed the |0.8| threshold for harmful multicollinearity, as is touched upon in section 7.1.

8.1.2 Legal status

When it comes to legal status, our finding that private limited liability companies are more likely to have access to finance than partnerships[30] are is consistent with findings for SSA (Beck & Cull, 2014, p. 595) and the EAC (Kira, 2013a, p. 60). We can however not conclude that being a private limited liability company has a greater effect than does being a sole proprietorship, as the confidence intervals for the two point estimates overlap. However, the point estimate for $LS\_private$ being of greater magnitude can serve as an indication that this might be the case, which would then be in line with the findings by Kira (2013a, p. 60).

The fact that $\Pr(CCS\_bin = 1)$ is estimated to be greater for sole proprietorships than for partnerships, is interesting since the sole proprietorship is considered the simplest form of business ownership. This implies that within the EAC, there is not a clear linear relationship between the degree of formality and the degree of access to finance. The reason for this could be that sole proprietorships, as opposed to partnerships which are typically bigger than sole proprietorships, are already subject to policies aimed at spurring the growth of very small businesses. An additional explanation can be found by looking to how the response variable is defined. $CCS\_bin = 1$ implies that the company has either had its loan application approved, *or has refrained from applying due to already having enough funding.* Interpreting the causal effect of being a sole proprietorship, instead of a partnership, as an increased probability of securing external credit, hinges on the company's growth ambitions being equivalent under both company forms. It is not unreasonable to instead assume sole proprietorships will consist f.e. of more small family-run businesses than partnerships will, and that these will be more content with keeping their company size relatively constant. In contrast, partnerships would then be expected to require more funding to finance future growth. As such, we would expect that the causal effect of being a sole proprietorship, as opposed to a partnership, is not an increased supply of funding, but a lowered demand for funding. This would then explain the seemingly contradictory effect we estimate.

8.1.3 Manager Experience and endogeneity

The estimated negative effect of having a manager with more experience of working in the sector of the firm, is somewhat counter-intuitive. Quartey et al. (2017, p. 24-25) also

[30]Note that the benchmark group consists of partnerships, as well as limited partnerships.

find significant effects of managerial experience, but these are estimated to be positive in Senegal and negative in Gambia. If we further examine research from another part of SSA, the findings by Fatoki and Asah (2011, p. 172-4) underline our puzzling result. They show a positive effect in South Africa of managerial *competence*, defined to include working experience. Kira (2013b, p. 4) also shows that in Tanzania, more experienced managers (> 5 years' experience) are more likely to get access to debt finance.

This begs the question of why our results differ so wildly. Apart from Quartey et al. (2017, p. 18), these other studies do not use the WBES datasets. When it comes to the response variable, all three of the other studies again differ from ours. The major difference is that Quartey et al. (2017) use a response variable based on the proportion of working capital financed with internal resources, whilst the other studies' response variables focus on access to debt financing. Our response variable stands out with its slightly more complex definition, as discussed in section 5.3.1. Perhaps the differences in outcomes simply stem from these differences in methodology.

If we return to the concept of managerial competence, used by Kira (2013b, p. 19) and Fatoki and Asah (2011, p. 172-173), and which includes the manager's level of education, one hypothesis could be that excluding education contributes to the counter-intuitive estimated effects of managerial experience. If education were to play an important role, then our model specification would be treating it as an omitted variable, which would instead be included in the error term ε. Based on the discussion in section 6.1 and appendix B, and using the same notation, we can determine what would be required for there to be a negative bias. Making the reasonable assumption that $\hat{\beta}_{education} > 0$, we would need to also assume that $\tilde{\delta}_{lmang\_exp} < 0$ for there to be a negative omitted variable bias. Even then, this is not expected to result in a large enough bias for the sign of $\hat{\beta}_{experience}$ to change. As such, we cannot reasonably attribute the estimated negative effect of $lmang\_exp$ to having omitted the effect of education.

On the topic of omitted variable bias, one of the main limitations of the WBES dataset is that it does not include a proper company performance measure. As such, firm performance is also treated as an omitted variable, from which bias can potentially arise. It is reasonable to assume that $\hat{\beta}_{performance} > 0$[31]. Hence, there could be negative bias if $\tilde{\delta}_{lmang\_exp} < 0$[32]. Hamori and Koyuncu (2015, p. 38) find evidence suggesting that this is the case. As a result, it is likely that the estimated negative effect of manager

[31] This can be interpreted as performance having a positive effect on CCS_bin.

[32] This can be interpreted as there being a negative covariance between performance and manager experience.

experience is due to not including a measure of company performance in the regression specification.

8.2 Country level

8.2.1 The results of the tests of overall significance

The fact that the χ^2-tests of overall significance do not reject for Burundi and Rwanda can plausibly be explained by there being few firms in the corresponding sub-samples, i.e. 103 and 156, respectively. However, the South Sudan sub-sample has 533 firms, compared to the Tanzania sub-sample of 253 firms, yet still does not even establish overall significance. It could be that when the South Sudan sub-sample was collected, it turned out non-representative of the country, purely by chance. Another explanation could be that the model is over-specified, and that the true model would include fewer explanatory variables.

Another more plausible explanation is that the financing environment in South Sudan and Burundi are substantially different from those of the other EAC countries, and that access to finance is not driven by the same kinds of firm characteristics in those countries. One source of indicators to look at is the *Worldwide Governance Indicators*, an ongoing research project by the World Bank that scores countries based on six dimensions of governance, including regulatory quality, rule of law, and voice and accountability (Kaufmann et al., 2011, p. 3-4). For example, South Sudan and Burundi respectively rank as the 4$^{\text{th}}$ and 13$^{\text{th}}$ worst[33] in the world in terms of rule of law; 1$^{\text{st}}$ and 18$^{\text{th}}$ worst in terms of government effectiveness; and 4$^{\text{th}}$ and 12$^{\text{th}}$ worst in terms of voice and accountability, all significantly lower[34] rankings than those of the other EAC member nations. One can also look at the 2020 *Ease of doing business* scores, released by the World Bank. These are country-wide aggregate measures of 12 areas of business regulation, including ease of starting a business, getting credit, and enforcing contracts (World Bank, 2019a, p.2, 19). These scores rank South Sudan and Burundi as the 6$^{\text{th}}$ and 26$^{\text{th}}$ worst in the world, respectively (World Bank, 2019a, p. 4). These rankings are also lower[35] than those of the rest of the EAC member nations, indicating a relatively non-business-oriented, and hence possibly substantially different, regulatory environment. Based on this and the WGI in-

[33] These rankings are all based on the 2018 WGI data.

[34] The WGI data reports point estimates of governance in standard normal units ranging between 2.5 and -2.5, as well as standard errors. Point estimates that differ by >1.96 standard errors are interpreted as significantly different with a confidence level of 95%.

[35] And in the case of Rwanda and Kenya, far lower.

dicators, it is not surprising that we do not find the same patterns within Burundi and South Sudan as in the other EAC countries.

8.2.2 Asymmetric information

Tanzania

Based on the estimated effects of the *lsize* variables, we can assert that there is some evidence of information asymmetry having an adverse effect on access to finance for SMEs in Tanzania, and that lessened information asymmetry would improve their prospects of obtaining financing. However, the effect of *lage* should be interpreted with caution due to it being excluded after controlling the FDR. These findings are somewhat in line with the findings by Kira and He (2012, p. 111-2, 115).

Kenya

In Kenya, the outcome of the *lage* variable suggests that reducing information asymmetry has a positive effect on access to finance, even though no significant effects of the other proxies for asymmetric information are estimated.

This can be connected to Kenya's success when it comes to boosting the financial inclusion in the country. One reason for the success is the increased access to traditional banking channels. Alper et al. (2018, p. 5) write that "Enterprise Surveys in 2007 and 2013 show that almost all firms have a checking or savings account, a bank loan, a line of credit, or overdraft facilities". It is likely that this has enabled companies to build long term relationships with banks, that contribute over time to reducing the degree of information asymmetry between the two. As a result, it is not surprising that increased firm age is shown to have a highly significant effect.

Uganda

For Uganda, like Kenya, only one proxy for information asymmetry is estimated to be significant - that being *lsize*. However, it is only at the 10% significance level and is not included after controlling the FDR. Even though evidence of information asymmetry having an effect on access to finance is found elsewhere in the study, as discussed for example in section 8.1, we cannot say that specifically Ugandan SMEs' access to finance

is significantly affected by information asymmetry[36].

8.2.3 Managerial Experience

An interesting outcome for all three countries is that just like in the sub-regional regression, managerial experience is estimated to have a negative effect. Although these are at varying significance levels and do not pass the FDR threshold, the point estimates all have the same sign as the sub-regional estimate, again underlining these seemingly counter-intuitive findings previously discussed. For further discussion about this, see section 8.1.3.

8.2.4 Legal Status

On the country level, legal status can only reasonably be said to have an effect in Kenya, but making such an interpretation would also cause us to expect $> 10\%$ of the findings to be erroneous, therefore it should be interpreted with some caution.

8.2.5 Foreign ownership

Uganda is the only country for which a significant effect of foreign ownership is estimated. However, this should also be interpreted with care, due to the effect being excluded upon controlling the FDR. This being said, finding positive effects of foreign ownership is not unprecedented in SSA, as Quartey et al. (2017, p. 23) find that this is the case within ECOWAS as well. Kira (2013a, p. 60, 63) also find similar effects within the EAC, when it comes to privately owned firms.

[36]The reader should not interpret this as concluding that no such effect exists in Uganda - only that we cannot claim with enough confidence that it does.

9. Conclusions

On the sub-regional level, access to finance is found to be significantly affected by company size, legal status and manager experience. Due to the estimated positive effect of size, we conclude that reducing information asymmetry has a positive effect on access to finance for SMEs in the EAC as a whole. As for legal status, we can conclude that being a private limited liability company is preferable to being any form of partnership, in terms of obtaining financing. However, we cannot establish that there is a linear relationship between the degree of formality and the degree of access to finance, as our results imply that the effect of being a sole proprietorship, as opposed to a partnership, is an increase in access to finance. We hypothesize that this is due to a combination of the definition of our response variable and the growth ambitions of sole proprietorships relative to partnerships. The estimated negative effect of manager experience is concluded to be a likely result of omitting a variable describing company performance.

On the country level, we can only establish for Kenya, Tanzania and Uganda that at least one covariate per country has an effect significantly different from 0. Because of the estimated significant positive effects of *lage* in Kenya and *lsize* in Tanzania, it can be concluded that reducing information asymmetry in these countries has a positive effect on access to finance.

9.1 Policy suggestions

All of the following policy suggestions are directed towards the EAC leadership (East African Community, 2020b) as the main implementer of the proposed systems, with each respective member nation as a participant that adjusts the policies to its unique circumstances. These policy suggestions are mainly aimed at reducing the degree of information asymmetry associated with SMEs, in order to achieve the UN Sustainable Development Goal previously mentioned (see sections 1 and 2.2), but also touch upon our results with regard to legal status and manager experience.

Education: SME managers need to become more knowledgeable about what they need to do to be granted a loan. Therefore, an educational scheme should be set up for the region. Based on the important characteristics for being granted a loan, mentioned in section 2.1, the program should include classes on how to create a business plan and project monthly cash flows. In addition, there should be classes that enhance bookkeeping

skills and provide knowledge about the importance of keeping a good relationship with a bank. Furthermore, since being a limited liability company is found to be preferable to being a partnership, classes on how to set up and run limited liability companies should also be included.

Enhance the credit information system and implement an SME database: Based on the estimated positive effect of reducing the degree of information asymmetry, and the benefits of a credit information system (see section 2.1), an appropriate step would be to expand the system in the region and to compile the information obtained, along with financial statements, in a dedicated SME database for the EAC region. This is to make such information accessible to financial institutions and governments. The intention is to help financial institutions assess the creditworthiness of the firms and to help governments get usable information to base policy decisions on. Inspiration can be obtained from the database of SMEs in Japan, as discussed in section 1.

Firm Size: As size has been found to have a positive significant effect in both the sub-regional regression and in the Tanzania regression, we recommend the implementation of policies that make it easier for small companies to grow larger. Two main obstacles to firm growth are the requirements of collateral and a repayment track record at the bank (see section 2.1). We suggest that the EAC as a region, or the respective governments, provides a form of repayment guarantee for SMEs wanting to get their first loan, provided the loan amount does not exceed some predetermined threshold. This way, they could get a chance to build up a track record of being able to stick to a repayment plan, hence increasing the likelihood of obtaining another loan and growing their business.

Manager Experience: Due to the conclusion that the significant negative effect of manager experience is most likely a result of omitted variable bias, we would recommend further investigations focusing on clarifying this effect. Until then however, polices aiming to provide SMEs with more experienced managers should be avoided, because they may end up having an *adverse* effect on access to finance.

9.2 Suggestions for future research

For future studies, we recommend incorporating a more robust measure of information asymmetry, perhaps an aggregate measure of *lage*, *size* and *audit*. This is to avoid issues with one proxy potentially being more valid than another, as well as to avoid interpretive difficulties that arise when only one proxy shows a significant effect.

Since we hypothesize that the low number of observations for Rwanda is the reason for the corresponding χ^2-test not rejecting, we suggest that future researchers attempt to include a larger sample for that country. An alternative would be to create a new survey dataset in a similar way to the one created by Kira and He (2012, p. 112-3).

Bibliography

AfDB. (2012). *Leveraging Capital markets for SME Financing in Rwanda: Draft Report*. African Development Bank Group. (Cit. on p. 4).

Akerlof, G. A. (1970). The market for "lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500 (cit. on p. 9).

Alper, E., Clements, B., Hobdari, N., Liu, L., Maina, J., Morales, A., Porcel, M. R., Perry, V., Saito, M., Shay, S., Tuitoek, K., & Weiss, J. (2018). *Kenya, Selected Issues* (No. 18/296). International Monetary fund. (Cit. on p. 26).

Aterido, R., Beck, T., & Iacovone, L. (2013). Access to Finance in Sub-Saharan Africa: Is There a Gender Gap? *World Development*, *47*, 102–120 (cit. on pp. 7, 39).

Beck, T., & Cull, R. (2014). SME finance in Africa. *Journal of African Economies*, *23*, 583–613 (cit. on pp. 1, 2, 7, 8, 23, 39).

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289–300 (cit. on pp. 5, 8, 16, 20).

Bigsten, A., Collier, P., Dercon, S., Fafchamps, M., Gauthier, B., Gunning, J. W., Oduro, A., Oostendorp, R., Patillo, C., Söderbom, M., Et al. (2003). Credit constraints in manufacturing enterprises in Africa. *Journal of African Economies*, *12*(1), 104–25 (cit. on p. 39).

Calice, P., Chando, V. M., & Sekioua, S. (2012). *Bank Financing to Small and Medium Enterprises in East Africa: Findings of a Survey in Kenya, Tanzania, Uganda and Zambia*. African Development Bank Group. (Cit. on pp. 1, 3, 9).

Chavez, E. (2017). *Credit Information and Firms' Access to Finance: Evidence from a New Credit-Constrained Status Measure*. The World Bank group. (Cit. on pp. 4, 9, 21, 39).

Cramer, J. S. (2003). *Logit Models from Economics and Other Fields*. Cambridge University Press. (Cit. on p. 44).

Cramer, J. S. (2005). *Omitted variables and misspecified disturbances in the logit model*. Tinbergen Institute Discussion Paper. (Cit. on p. 44).

Divakaran, S., J. McGinnis, P., & Shariff, M. (2014). *Private Equity and Venture Capital in SMEs in Developing Countries*. The World Bank Group. (Cit. on p. 4).

Diz, A. P., Carvajal-Rodrıéguez, A., & Skibinski, D. O. (2011). Multiple hypothesis testing in proteomics: a strategy for experimental work. *Molecular & Cellular Proteomics*, *10*(3) (cit. on p. 16).

East African Community. (2018). *Agriculture and Food Security*. Retrieved June 4, 2020, from https://www.eac.int/agriculture. (Cit. on p. 10)

East African Community. (2012). *EAC Industrial Upgrading and Modernisation Programme Overview*. Retrieved May 27, 2020, from https://www.eac.int/component/content/article/123-political-federation/204-laying-foundation.html. (Cit. on p. 2)

East African Community. (2020a). *EAC Partner States*. Retrieved April 10, 2020, from https://www.eac.int/eac-partner-states. (Cit. on p. 10)

East African Community. (2020b). *Leadership Team*. Retrieved May 27, 2020, from https://www.eac.int/about-eac/leadership. (Cit. on p. 28)

European Union. (2003). Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. *Official Journal*, *L 124*, 30–47 (cit. on p. 1).

Fatoki, O., & Asah, F. (2011). The Impact of Firm and Entrepreneurial Characteristics on Access to Debt Finance by SMEs in King Williams' Town, South Africa. *International Journal of Business and Management*, *6* (cit. on pp. 24, 39).

Fidrmuc, J., Schreiber, P., & Siddiqui, M. (2015). Information Asymmetry, Relationship Banking and Financing Costs of SMEs. *SSRN Electronic Journal* (cit. on p. 9).

Fjose, S., Grünfeld, L. A., & Green, C. (2010). SMEs and growth in Sub-Saharan Africa: Identifying SME roles and obstacles to SME growth. *MENON Business Economics Publication*, (14) (cit. on p. 1).

Fombang, M., & Adjasi, C. (2018). Access to finance and firm innovation. *Journal of Financial Economic Policy*, *10*, 73–95 (cit. on pp. 2, 7, 8).

Fowowe, B. (2017). Access to finance and firm performance: Evidence from African countries. *Review of Development Finance*, *7*(1), 6–17 (cit. on pp. 2, 3, 8, 11, 12).

Gertler, M. (1988). Financial Structure and Aggregate Economic Activity: An Overview. *Journal of Money, Credit and Banking*, *20*(3), 559–88 (cit. on p. 9).

Hamori, M., & Koyuncu, B. (2015). Experience matters? The impact of prior CEO experience on firm performance. *Human Resource Management*, *54*(1), 23–44 (cit. on p. 24).

IMF. (2018). *Sub-Saharan Africa Regional Economic Outlook: Domestic Revenue Mobilization and Private Investment*. International Monetary Fund. (Cit. on p. 2).

Kalash, I., & Sabsabi, M. (2018). The Effect of Information Asymmetry on the Relationship Between Leverage and the Performance of the Turkish Industrial Firms, In *International Symposium on Economics, Politics and Administration*. (Cit. on p. 9).

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: Methodology and analytical issues. *Hague Journal on the Rule of Law*, *3*(2), 220–246. Full data available at http://info.worldbank.org/governance/wgi/ (cit. on p. 25).

Kennedy, P. (2008). *A Guide to Econometrics* (6th ed.). Wiley. (Cit. on p. 16).

Kira, A. R. (2013a). Determinants of financing constraints in East African countries' SMEs. *International Journal of Business and management*, *8*(8), 49 (cit. on pp. 1, 3, 5, 7, 9, 21–23, 27, 39).

Kira, A. R. (2013b). The Evaluation of the Factors Influence the Access to Debt Financing by Tanzanian SMEs. *European Journal of Business and Management*, *5*, 1–24 (cit. on pp. 24, 39).

Kira, A. R., & He, Z. (2012). The Impact of Firm Characteristics in Access of Financing by Small and Medium-sized Enterprises in Tanzania. *International Journal of Business and Management*, *7* (cit. on pp. 26, 30).

Kuntchev, V., Ramalho, R., Rodrıéguez-Meza, J., & Yang, J. S. (2012). What have we learned from the Enterprise Surveys regarding access to finance by SMEs. *Enterprise Analysis Unit of the Finance and Private Sector Development, The World Bank Group* (cit. on pp. 3, 7, 8, 10–12, 21, 39, 40).

Kuwahara, S., & Yoshino, N. (2015). *Role of the Credit Risk Database in Developing SMEs in Japan: Lessons for the Rest of Asia*. Asian Development Bank Institute. (Cit. on p. 1).

Lee, N., Sameen, H., & Cowling, M. (2015). Access to finance for innovative SMEs since the financial crisis. *Research policy*, *44*(2), 370–80 (cit. on p. 39).

Ministry of Trade, Industry and Cooperatives. (2015). *Uganda Micro, Small and Medium Enterprises (MSME) Policy*. Ministry of trade, Industry and Cooperatives - The Republic of Uganda. (Cit. on pp. 4, 11, 42).

Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*, *26*(1), 67–82 (cit. on p. 44).

Mugwika, G. (2013). The Impact of the East African Community Integration on Small & Medium Enterprises. Case of Kenya. *Academy of Management Proceedings*, *1*(1) (cit. on p. 3).

Mundfrom, D., Perrett, J., Schaffer, J., Piccone, A., Roozeboom, M., Et al. (2006). Bonferroni adjustments in tests for regression coefficients. *Multiple Linear Regression Viewpoints*, *32*, 1–6 (cit. on pp. 8, 16).

Quartey, P., Turkson, E., Abor, J. Y., & Iddrisu, A. M. (2017). Financing the growth of SMEs in Africa: What are the contraints to SME financing within ECOWAS? *Review of Development Finance*, *7*(1), 18–28 (cit. on pp. 1, 3, 7, 9, 21–24, 27, 39).

Republic of Kenya. (2015). The Companies Act. *Kenya Gazette Supplement*, *158*(17) (cit. on p. 22).

Sacerdoti, E. (2005). *Access to Bank Credit in Sub-Saharan Africa: Key Issues and Reform Strategies* (No. 05/166). International Monetary Fund. (Cit. on p. 10).

Söderbom, M. (2009a). Applied Econometrics, Lecture notes: Binary Choice Models. Department of Economics at the University of Gothenburg. Retrieved May 3, 2020, from https://www.soderbom.net/lecture10notes.pdf. (Cit. on pp. 8, 13, 14, 43)

Söderbom, M. (2009b). Applied Econometrics, Lecture notes: Nonlinear Models Beyond Binary Choice: Multinomial Response Models, Corner Solution Models & Censored Regressions. Department of Economics at the University of Gothenburg. Retrieved May 4, 2020, from https://www.soderbom.net/lecture13notesfinal.pdf. (Cit. on p. 11)

Streiner, D. L. (2015). Best (but oft-forgotten) practices: the multiple problems of multiplicity—whether and how to correct for many statistical tests. *The American journal of clinical nutrition*, *102*(4), 721–728 (cit. on pp. 8, 16).

Thomas, S. (2002). Firm diversification and asymmetric information: Evidence from analysts' forecasts and earnings announcements. *Journal of Financial Economics*, *64*(3), 373–396 (cit. on p. 9).

UN. (2018). *Promote inclusive and sustainable economic growth, employment and decent work for all*. Retrieved April 27, 2020, from https://www.un.org/sustainabledevelopment/economic-growth/. (Cit. on p. 1)

Vandenberg, P., Chantapacdepong, P., & Yoshino, N. (2016). *SMEs in Developing Asia New Approaches to Overcoming Market Failures*. The Asian Development Bank. (Cit. on p. 1).

Werner, S., Malaspina, D., & Rabinowitz, J. (2007). Socioeconomic status at birth is associated with risk of schizophrenia: Population-based multilevel study. *Schizophrenia bulletin*, *33*(6), 1373–1378 (cit. on p. 15).

Wooldridge, J. (2015). *Introductory Econometrics: A Modern Approach*. Cengage Learning. (Cit. on pp. 8, 13–15, 44).

World Bank. (2019a). *Doing Business 2020*. The World Bank Group. (Cit. on p. 25).

World Bank. (2014). *MSME country indicators 2014*. Retrieved April 28, 2020, from https://finances.worldbank.org/widgets/psn8-56xf. (Cit. on pp. 4, 11, 42)

World Bank. (2019b). *Rwanda Systematic Country Diagnostic*. The World Bank Group. (Cit. on p. 2).

World Bank. (2017). *United Republic of Tanzania Systematic Country Diagnostic*. The World Bank Group. (Cit. on p. 2).

A. Tables and figures appendix

Abbreviation	Meaning
EAC	East African Community
ECOWAS	Economic Community of West African States
FDR	False discovery rate
OLS	Ordinary least squares
SME	Small- and medium-sized enterprise
MSME	Micro-, small- and medium-sized enterprise
SSA	Sub-Saharan Africa
VIF	Variance inflation factor
WBES	World Bank Enterprise Survey
WGI	Worldwide Governance Indicator

Table A.1: Abbreviations

Notation	Meaning
Subscript i	Denotes observation
Subscript j	Denotes regressor
$\Leftrightarrow$	"If and only if"
Hat operator $\hat{}$	Denotes estimated parameter
$\in$	"Is in the set"
$\notin$	"Is not in the set"
log()	Natural logarithm operator, for consistency with other literature
$[z]$	Iverson brackets, denoting a variable taking on the value 1 if the condition z is fulfilled, and 0 otherwise
$\Pr(z)$	Probability operator, denoting the probability of event z occuring
$\exp(z)$	Exponentiation operator, denoting e^z
P	$\Pr(CCS\_bin = 1)$
$G(z)$	Denotes a function of z, without specifying the actual function. Used as a placeholder when multiple different functions could be substituted in
$\Lambda(z)$	Denotes the logistic function $\frac{\exp(z)}{1+\exp(z)}$
$E[z]$	Expected value operator

Table A.2: Shared notation

				CCS_bin	
Country	**Year**	**Freq.**	**Percentage**	**Mean**	**STDEV**
Burundi	2014	103	5.02	0.56	0.50
Kenya	2018	633	30.88	0.66	0.47
Rwanda	2011	156	7.61	0.60	0.49
South Sudan	2014	533	26.00	0.39	0.49
Tanzania	2013	253	12.34	0.35	0.48
Uganda	2013	372	18.15	0.47	0.50
Total		2,050	100		

Table A.3: Summary of companies in the sample

Variable	**Obs.**	**Mean**	**STDEV**	**Min**	**Max**
lage	2,050	2.20	0.97	0	4.82
lsize	2,050	2.56	0.81	1.61	4.61
lmang_exp	2,050	2.26	0.79	0	4.17
mang_fem	2,050	0.14	0.35	0	1
exporter	2,050	0.06	0.23	0	1
for_owner	2,050	0.18	0.38	0	1
audit	2,050	0.52	0.50	0	1
LS_private	2,050	0.15	0.36	0	1
LS_sole	2,050	0.49	0.50	0	1
lgCity	2,050	0.57	0.49	0	1

Table A.4: Variable descriptive statistics

	EAC	Burundi	Kenya	Rwanda	South Sudan	Tanzania	Uganda
lage	2.11	1.50	1.55	1.34	1.21	1.35	1.30
lsize	1.34	1.23	1.19	1.20	1.35	1.60	1.19
lmang_exp	1.54	1.31	1.56	1.32	1.27	1.30	1.29
mang_fem	1.02	1.19	1.05	1.13	1.02	1.03	1.04
exporter	1.06	1.09	1.07	1.31	1.03	1.09	1.11
for_owner	1.27	1.16	1.05	1.32	1.34	1.11	1.18
audit	1.40	1.04	1.17	1.40	1.18	1.40	1.21
LS_private	1.26	2.82	1.19	1.29	1.18	2.25	1.10
LS_sole	1.52	3.02	1.30	1.36	1.34	2.08	1.39
lgCity	1.23	1.44	1.16	1.26	1.37	1.18	1.13

Table A.5: VIF results

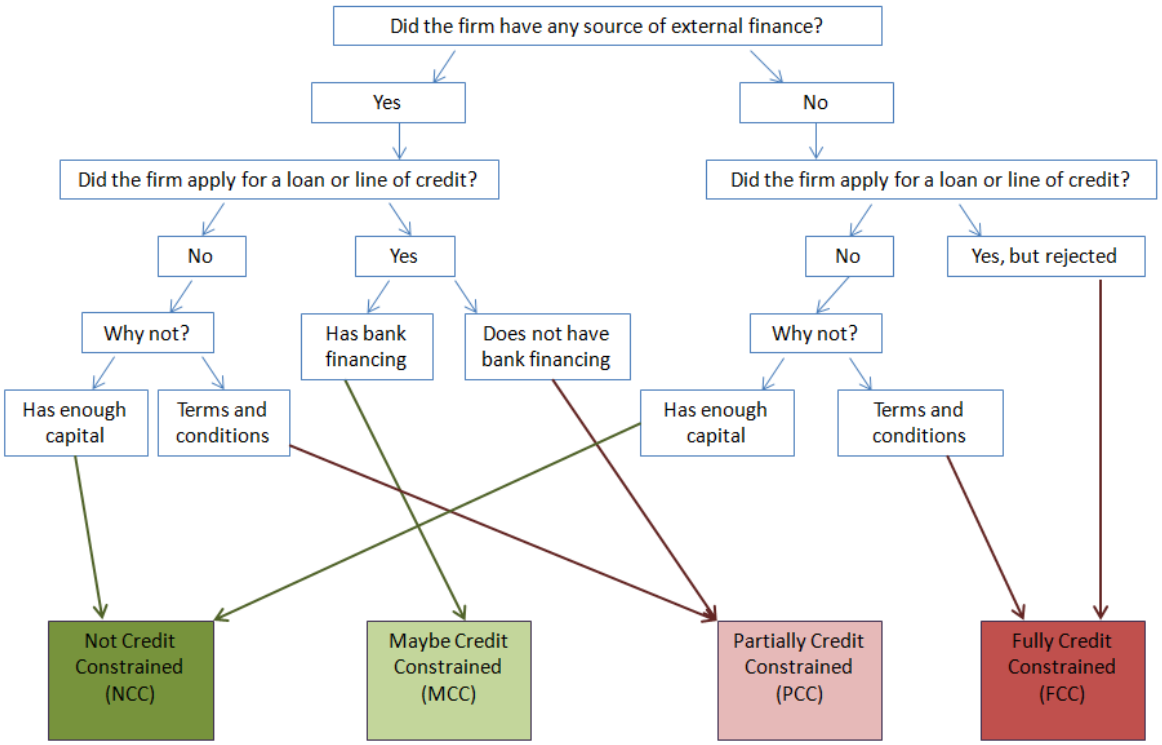
	lage	*lsize*	*lmang_exp*	*mang_fem*	*exporter*	*for_owner*	*audit*	*LS_private*	*LS_sole*	*lgCity*
lage	1.000									
lsize	0.316	1.000								
lmang_exp	0.558	0.240	1.000							
mang_fem	0.042	−0.055	−0.037	1.000						
exporter	0.104	0.167	0.068	0.001	1.000					
for_owner	−0.223	0.065	−0.088	−0.026	0.076	1.000				
audit	0.285	0.353	0.222	−0.002	0.137	0.016	1.000			
LS_private	0.137	0.202	0.053	0.000	0.081	0.075	0.207	1.000		
LS_sole	−0.043	−0.290	−0.099	−0.009	−0.123	−0.258	−0.306	−0.415	1.000	
lgCity	−0.003	0.162	0.189	−0.008	0.046	0.128	0.112	0.037	−0.139	1.000

Table A.6: Explanatory variable correlations

Explanatory Variable	**Description**	**Used in studies**
lage	Natural logarithm of the company's age in years	Aterido et al. (2013); Beck and Cull (2014); Bigsten et al. (2003); Chavez (2017); Fatoki and Asah (2011); Kira (2013a); Kira (2013b); Kuntchev et al. (2012); Quartey et al. (2017)
lsize	Natural logarithm of the number of permanent, full-time employees last fiscal year	Aterido et al. (2013); Beck and Cull (2014); Bigsten et al. (2003); Chavez (2017); Fatoki and Asah (2011); Kira (2013a); Kira (2013b); Kuntchev et al. (2012); Lee et al. (2015); Quartey et al. (2017)
lmang_exp	Natural logarithm of the top manager's years of experience of working in the company's sector	Fatoki and Asah (2011); Kira (2013b); Quartey et al. (2017)
mang_fem	Dummy $=1 \Leftrightarrow$ the manager is female	Aterido et al. (2013); Beck and Cull (2014); Chavez (2017); Kuntchev et al. (2012); Lee et al. (2015); Quartey et al. (2017)
exporter	Dummy $=1 \Leftrightarrow \geq 10\%$ of sales are direct exports	Aterido et al. (2013); Kuntchev et al. (2012); Quartey et al. (2017)
for_owner	Dummy $=1 \Leftrightarrow \geq 10\%$ of the company is owned by foreigners	Aterido et al. (2013); Beck and Cull (2014); Bigsten et al. (2003); Chavez (2017); Kira (2013a); Kuntchev et al. (2012); Quartey et al. (2017)
audit	Dummy $=1 \Leftrightarrow$ the company's financial statements have been audited	Chavez (2017)
LS_private	Dummy $=1 \Leftrightarrow$ the company is a private LLC	Aterido et al. (2013); Beck and Cull (2014); Bigsten et al. (2003); Fatoki and Asah (2011); Kira (2013a); Kira (2013b)
LS_sole	Dummy $=1 \Leftrightarrow$ the company is a sole proprietorship	Aterido et al. (2013); Beck and Cull (2014); Bigsten et al. (2003); Fatoki and Asah (2011); Kira (2013a); Kira (2013b)
lgCity	Dummy $=1 \Leftrightarrow$ the company is located in the capital city or a city with $\geq 1M$ inhabitants	Aterido et al. (2013); Bigsten et al. (2003)

This table also lists studies that use either non-logged values or dummy variables for age and size.

Table A.7: Explanatory variable definitions



Source: Kuntchev et al. (2012, p. 20)

Figure A.1: Simplified definition of the groups comprising CCS

Dependent Variable							
CCS_bin	EAC	Burundi	Kenya	Rwanda	South Sudan	Tanzania	Uganda
lage	0.036**	−0.078	0.106***	−0.009	0.003	0.086	−0.033
	(0.016)	(0.068)	(0.026)	(0.052)	(0.031)	(0.054)	(0.042)
lsize	0.071***	0.167***	0.031	−0.039	0.081*	0.190***	0.067*
	(0.015)	(0.064)	(0.023)	(0.047)	(0.042)	(0.036)	(0.037)
lmang_exp	−0.043**	−0.056	−0.053*	0.012	−0.008	−0.080	−0.104**
	(0.017)	(0.070)	(0.029)	(0.062)	(0.031)	(0.050)	(0.044)
mang_fem	0.025	−0.025	0.019	−0.051	0.027	0.097	0.082
	(0.031)	(0.131)	(0.051)	(0.098)	(0.070)	(0.093)	(0.066)
exporter	0.078	−0.184	0.065	−0.127	0.202	−0.082	0.161
	(0.048)	(0.179)	(0.063)	(0.211)	(0.159)	(0.126)	(0.121)
for_owner	−0.014	0.035	−0.084	0.162	−0.048	0.164	0.186**
	(0.031)	(0.160)	(0.060)	(0.111)	(0.048)	(0.142)	(0.089)
audit	0.020	0.025	0.079	0.200**	−0.020	−0.111*	−0.010
	(0.025)	(0.095)	(0.049)	(0.084)	(0.050)	(0.065)	(0.055)
LS_private	0.086***	−0.082	0.134***	0.114	−0.089	0.173*	0.225
	(0.033)	(0.210)	(0.051)	(0.110)	(0.069)	(0.101)	(0.166)
LS_sole	0.062**	−0.150	0.045	0.260***	0.025	0.182**	0.092
	(0.026)	(0.189)	(0.045)	(0.083)	(0.048)	(0.081)	(0.064)
lgCity	−0.001	0.030	−0.022	−0.132	−0.064	0.124	0.010
	(0.024)	(0.113)	(0.040)	(0.283)	(0.048)	(0.078)	(0.053)
Country fixed effects	Yes	No	No	No	No	No	No
N	2050	103	633	156	533	253	372
pseudo R^2	0.0614	0.0842	0.0506	0.0662	0.0158	0.1269	0.0493
chi2	169.25***	13.94	37.73***	14.14	10.50	39.12***	26.25***

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Reported values are APEs, as opposed to coefficients

Table A.8: Probit regression results

Country	**Micro**	**Small**	**Medium**	**Large**	**Data Year**
Rwanda	1-3	4-30	31-100	>100	2011
South Sudan	1-4	5-49	50-99	$\geq$ 100	2012
Burundi	<25	<50	<250	$\geq$ 250	2010
Kenya*	1-10	11-50	51-100	>100	2012
Uganda*	1-4	5-50	50-100	>100	2012
Tanzania	1-4	5-49	50-99	>100	2012
Median upper limit	4	49.5	100	100	
Range used	1-4	5-50	51-100	>100	

Table A.9: Definitions of MSMEs in the EAC

Source: World Bank (2014) (* are from Ministry of Trade, Industry and Cooperatives (2015, p. 25))

B. Logit appendix

Having defined the latent variable y^*, and following the methodology by Söderbom (2009a, p. 11), we also make the assumption that ε follows a logistic distribution, such that $\Lambda(\varepsilon) = \frac{\exp(\varepsilon)}{1+\exp(\varepsilon)}$. Next, it naturally follows that

$$\begin{aligned}
\Pr(CCS_bin = 1|\boldsymbol{x}) =& \Pr(y^* > 0|\boldsymbol{x}) \\
=& \Pr(\boldsymbol{x\beta} + \varepsilon > 0|\boldsymbol{x}) \\
=& \Pr(\varepsilon > -\boldsymbol{x\beta}|\boldsymbol{x}) \\
=& 1 - \Lambda(-\boldsymbol{x\beta}) \\
=& \Lambda(\boldsymbol{x\beta}),
\end{aligned} \tag{B.1}$$

where in the final steps, we exploit the symmetry of the logistic function. Consequently, the appropriate model to use for finding determinants of $\Pr(CCS\_bin = 1|\boldsymbol{x})$, is the logistic regression model, such that $\Pr(CCS\_bin = 1|\boldsymbol{x}) = \Lambda(\boldsymbol{x\beta})$. This can then be rewritten as

$$\begin{aligned}
P =& \Lambda(\boldsymbol{x\beta}) \\
P =& \frac{\exp(\boldsymbol{x\beta})}{1 + \exp(\boldsymbol{x\beta})} \\
P + Pe^{\boldsymbol{x\beta}} =& e^{\boldsymbol{x\beta}} \\
P =& e^{\boldsymbol{x\beta}} - Pe^{\boldsymbol{x\beta}} \\
P =& e^{\boldsymbol{x\beta}}(1 - P) \\
\frac{P}{1-P} =& e^{\boldsymbol{x\beta}} \\
\log\left(\frac{P}{1-P}\right) =& \boldsymbol{x\beta}
\end{aligned} \tag{B.2}$$

to allow for interpretation of the coefficients. Since the odds that y=1 can be rewritten as $\frac{\Pr(y=1)}{1-\Pr(y=1)}$ through

$$Odds(y = 1) = \frac{\sum_{i=1}^{n}[y = 1]}{\sum_{i=1}^{n}[y = 0]} = \frac{n^{-1}\sum_{i=1}^{n}[y = 1]}{n^{-1}\sum_{i=1}^{n}[y = 0]} = \frac{\Pr(y = 1)}{1 - \Pr(y = 1)}, \tag{B.3}$$

$\log\left(\frac{P}{1-P}\right)$ is the same as log(odds).

As outlined by Wooldridge (2015, p. 527), the marginal effect[37] of x_j, can be obtained by solving the partial derivative[38]

$$\frac{\partial P}{\partial x_j} = \frac{d\Lambda(\boldsymbol{x\beta})}{d(\boldsymbol{x\beta})}\beta_j, \text{ where } \Lambda(\boldsymbol{x\beta}) = \frac{\exp(\boldsymbol{x\beta})}{1+\exp(\boldsymbol{x\beta})}, \tag{B.4}$$

meaning that the marginal effect of x_j depends on the levels of *all* regressors. However, this is only relevant for estimating the *magnitude* of the marginal effect. Since $\frac{d\Lambda(z)}{dz} > 0$ for all z, the marginal effect of x_j will have the same sign as $\hat{\beta}_j$ (Wooldridge, 2015, p. 527). One of many ways of quantifying the marginal effect of x_j is by averaging its marginal effect across all observations in the sample (Wooldridge, 2015, p. 532), computed as

$$n^{-1}\left(\sum_{i=1}^{n}\frac{d\Lambda(\boldsymbol{x_i}\hat{\boldsymbol{\beta}})}{d(\boldsymbol{x_i}\hat{\boldsymbol{\beta}})}\right)\hat{\beta}_j. \tag{B.5}$$

This measure is known as the average partial effect (APE)[39].

Omitted variable bias

The reader should be aware that in logit regression, omitted variable bias works in a slightly different way from how it works in OLS. For example, it can be shown that when excluding an explanatory variable x_k that is *uncorrelated* with each $x_1 \dots x_{k-1}$, each $\hat{\beta}_1 \dots \hat{\beta}_{k-1}$ gets depressed towards 0 (Mood, 2010, p. 68-9), while (importantly) APEs are left mostly unaffected (Cramer, 2005, p. 4-7). Also, it can be shown (Mood, 2010, p. 69; Cramer, 2003, p. 80-1) that excluding a *correlated* explanatory variable from a logit regression will cause us to estimate

$$\tilde{\beta}_j = (\hat{\beta}_j + \hat{\beta}_k\tilde{\delta}_j)\frac{\sqrt{3.29}}{\sqrt{3.29+\beta_k^2\sigma_v^2}}, \tag{B.6}$$

where σ_v^2 = the variance of the error term v_i when regressing x_k on each $x_1 \dots x_{k-1}$. This is similar to the OLS case, except that the biased coefficient is also scaled by the positive $\frac{\sqrt{3.29}}{\sqrt{3.29+\beta_k^2\sigma_v^2}}$ term. Importantly, since $\widehat{\text{APE}}_j$ always has the same sign as $\hat{\beta}_j$, then if $\frac{|\hat{\beta}_j|}{\hat{\beta}_j} \neq \frac{|\hat{\beta}_k\tilde{\delta}_j|}{\hat{\beta}_k\tilde{\delta}_j}$ and $|\hat{\beta}_j| < |\hat{\beta}_k\tilde{\delta}_j|$, then omitted variable bias will have caused both $\hat{\beta}_j$ and $\widehat{\text{APE}}_j$ to change sign.

[37] This is for a continuous x_j. For marginal effects of binary variables, see Wooldridge (2015, p. 527).
[38] It should be noted that the notation here is slightly altered from that of the original source.
[39] It is also sometimes referred to as the average marginal effect.