# MIND THE DATA GAPS: AN EXAMINATION OF WOMEN-OWNED ENTERPRISE REPRESENTATION

MOST RECENT DRAFT

Morgan Hardy
New York University Abu Dhabi

Gisella Kagy
Vassar College

Nusrat Jimi
Vassar College

February 14, 2023

**Abstract**

Using data from 43 countries in Sub-Saharan Africa, we document large variations in women-owned enterprise representation and estimates of gender gaps in enterprise performance between commonly available data sources. We provide empirical evidence that these differences are driven by variations in gender-blind sampling protocols. Women-owned enterprises are less likely to meet the sampling criteria for most widely available enterprise data and those that do are more positively selected on performance, relative to male-owned enterprises. We document differences in implied policy and research priorities; sources with higher women-owned enterprise representation point toward issues of market access, over more commonly studied barriers.

**JEL Classifications: D22, J16, J46, L26, O12**
**Word Count:** 5986

We are grateful to Kathleen Beagle, Gaurav Chiplunkar, Asif Mohammed Islam, David McKenzie, Jorge Rodriguez Meza, Aishwarya Ratan, our colleagues at The Ethiopian Economic Association, and attendees of the Regional Workshop on Global Foundational Analysis to Close Gender Profitability Gap in Addis Ababa, Ethiopia, for helpful comments and suggestions related to the broader project that has generated this paper. We are grateful to Denat Negatu and Juan Pablo Rossi for outstanding research assistance on both the broader project and this draft. The broader project is funded by a grant from The Bill and Melinda Gates Foundation administered by The Ethiopian Economic Association. Please e-mail morgan.hardy@nyu.edu, gikagy@vassar.edu, and njimi@vassar.edu with any questions, comments, or suggestions.

# 1 Introduction

Economics is increasingly an empirical science. Central to academic research and evidence-based policy-making is quality data. Over the last century, there has been an increased effort by policymakers, non-governmental organizations, and international institutions to compile usable and publicly available data that form the backbone of research agendas and policy directions alike. As our global society moves toward an increased focus on equity and inclusion, we naturally find ourselves using these existing data sources to pursue a better understanding of such issues. This paper highlights the importance of considering the original sampling frame in the secondary use of data for new research, by examining the case of women-owned enterprise representation.

There is growing momentum to improve women's economic empowerment and achieve gender parity more broadly in low-income countries (UNDP, 2015). A key component of a woman's economic empowerment is equity in employment opportunities and the relative compensation received. In Sub-Saharan Africa, employment opportunities for women are most commonly, and often exclusively, in self-employment (Gindling and Newhouse, 2014). In response to this fact, a growing body of work focuses on understanding and reducing the gender profitability gap faced by enterprise owners (Delecourt and Fitzpatrick, 2021; Hardy and Kagy, 2018,2; Nix, Gamberoni, and Heath, 2015).

Representative enterprise data is an important tool needed to push forward this endeavor. Some such data comes from independently collected micro-data sets that are representative of specific industries and contexts.[1] However, more broadly representative enterprise data is a key resource for generating stylized facts and documenting trends that drive policy and research agendas focused on dismantling barriers that are holding back female entrepreneurs.

We document a gender gap in representation within the most widely available enterprise data. Using data from 43 Sub-Saharan African countries, we compare the rates of female ownership in samples from two different sources: 1) the World Bank Enterprise Surveys (WBES) of which there are three types: the Enterprise Survey, Micro Enterprise Survey, and Informal Sector Enterprise Survey,[2]

[1] Examples include garment making enterprises in a specific Ghanaian town (Hardy and Kagy, 2018,2), microentrepreneurs in select areas of Uganda (Delecourt and Fitzpatrick, 2021); vegetable sellers in Jaipur India (Delecourt and Ng, 2021); microenterprises in Kampala, Uganda (Riley, 2020). The Profiting from Parity 2019 World Bank Report compiles a summary of findings from several micro-data sets (World Bank Group, 2019).

[2] The Enterprise Survey is the most commonly available and widely used source of enterprise data and focuses on formal establishments with 5 or more employees (Enterprise Analysis Unit, 2021). The Micro Enterprise Survey and the Informal

and 2) enterprises identified in nationally representative multi-topic household surveys (HHS) with modules on non-farm businesses. We observe relatively lower women-owned enterprise representation in the WBES samples relative to the enterprises identified in the HHS. This trend holds within data from the same country, the same year, and the same country and year.

We also detect economically and statistically significant differences in estimates of gender gaps in business performance. Using total annual sales data from the WBES and HHS enterprises, we estimate the female- to male-owned enterprise sales ratio at the data source level and then compare this ratio across data sources. We find that female- and male-owned businesses report almost equal sales in the most commonly collected Enterprise Survey and Micro Enterprise Survey. In contrast, HHS data yield an average female- to male-owned enterprise sales ratio of 0.59. This difference is not explained by differences in data source countries or years.

Why do we see such differences in representation and estimates of business performance gaps by owner gender? We provide empirical evidence that differences in the (gender-blind) sampling protocols between data sources drive this variation. Using the HHS data, we compare the female ownership percentages across groups of businesses that match the sampling frame characteristics of either the Micro Enterprise Survey or the Informal Sector Enterprise Survey, relative to those that do not.[3] We find that HHS enterprises that satisfy the sampling frame criteria of either the Micro Enterprise or Informal Sector Enterprise Survey are significantly less likely to have female owners compared to businesses that do not. The differences are not explained by the relatively urban focus of the Micro Enterprise and Informal Sector Enterprise Survey; rural businesses are no more likely to have female owners than urban businesses with similar other characteristics. Rather, we conclude that the persistent gaps in relative rates of female ownership across the WBES surveys and HHS is primarily due to their differences in the other sampling target characteristics. The Micro Enterprise Survey explicitly focuses on formal businesses and the Informal Sector Enterprise Survey de facto samples businesses with physical structures (i.e., operated outside of the home) by approaching visible clusters of urban enterprises.

Next, we empirically demonstrate how differential selection among women-owned enterprises leads to differences in the estimated female- to male-owned enterprise sales ratio across data

Sector Enterprise Survey focus on formal microenterprises and informal businesses in urban and metropolitan areas, respectively.

[3]There is a limited overlap of the Enterprise Survey target sampling characteristics with the average business found via the household. Only 0.29% of HHS enterprises meet the criteria, and therefore we are unable to conduct meaningful quantitative analysis including this group.

sources. We compare differences by gender in the total annual sales of HHS businesses across groups with characteristics that mirror the sampling protocol of either the Micro Enterprise Survey or the Informal Sector Enterprise Survey and those that do not. Women-owned enterprises that have similar characteristics to the businesses surveyed by either the Micro Enterprise Survey or Informal Sector Enterprise Survey are more positively selected on performance compared to their respective male-owned enterprise groups.

Finally, we demonstrate contrasting implications for research and policy priorities, using the self-reported enterprise constraints from different data sources to generate estimates of the owners' implied resource preferences. We find that the WBES surveys point toward a different set of enterprise-focused policy and research priorities than those highlighted by the HHS. Primarily, the WBES suggests that governance and safety issues are key enterprise barriers while the HHS suggests market access as the key constraint. We note that these different implications can be due to the differing preferences of those represented but also stem from differences in measurement (i.e., question structure) across survey types. By comparing the resource preferences implied from true responses with those implied from randomly generated survey responses, we find that differences across data sources are driven both by sampling protocols and survey structure. Overall, this exploration demonstrates the importance of considering both survey structure and sampling protocol in the policy implications drawn from data.

It is important to note that the findings of our paper do not imply that the WBES data is uniquely under-representing women nor that it is not useful for other purposes. In fact, we find very similar levels of women-owned enterprise representation when we compare the Ethiopian 2006 and 2015 Enterprise Survey data with the Ethiopian Manufacturing Census data (which focuses on formal manufacturing enterprises that engage ten or more persons and use power-driven machinery) for the same years. Even the Ghana 2014 Economic Census data, which essentially lists all businesses with a physical structure, yields lower women-owned enterprise representation than the 2013 Ghana HHS data.[4] We suspect this is due to the differential propensity of women-owned businesses to operate exclusively inside the household as discussed above.

The key takeaway from our analysis is that, like many pre-existing and widely available public data sets, neither the WBES Surveys nor the non-farm business modules in the HHS are explicitly designed to answer the questions about gender gaps that are now of increasing interest to researchers

[4]This finding echoes Kerr and McDougall (2020) that firm census coverage, both within and across countries, can vary dramatically and may not represent the average firm.

and policymakers. The WBES data (and other comparable data sources with similar sampling protocols) are often used to inform enterprise policy.[5] Our analysis makes clear that insights derived from these data sources may be gender-biased toward the average male owner's experience. A gender-aware sampling approach to data collected on enterprises is a key consideration in future data collection efforts focused on creating common data sets for research on gender gaps.

Systematic under-representation of certain demographic groups in scientific research is an issue that extends across disciplines: for example, there is a reliance on convenience samples of college students in behavioral science (Henrich, Heine, and Norenzayan, 2010), a common focus on men in economic history (Bailey, Anderson, and Massey, 2017; Jácome, Kuziemko, and Naidu, 2021), and large exclusion of 'vulnerable' populations in medical research (Michelman and Msall, 2021; Murthy, Krumholz, and Gross, 2004). Our paper contributes to this growing literature on unequal representation by highlighting the need for secondary users of data to carefully consider the data sampling frame and its implications for representation. An equity-aware sampling approach to public good data collection efforts is a key consideration as we move toward a more equitable evidence-based society.

## 2 Data

The bulk of analysis in this paper utilizes all publicly available WBES data and HHS data for Sub-Saharan Africa collected after 2005.[6] The WBES data includes 85 Enterprise Surveys covering 43 countries, 26 Micro Enterprise Surveys covering 24 countries, and 18 Informal Sector Enterprise Surveys covering 16 countries. The HHS data includes 39 Multi-topic Household Surveys covering 15 countries[7]. See Appendix Table A1 for a list of countries and years of data collection.

[5] Recent examples include (Abor and Quartey, 2010; Eifert, Gelb, and Ramachandran, 2008; Hallward-Driemeier and Pritchett, 2015), and (Fang, Goh, Roberts, Xu, and Zeufack, 2020).

[6] Although some WBES were conducted before 2006, there was considerable heterogeneity across countries in terms of the questionnaire format, sectors covered, and sampling methodology. To match the WBES timeframe, we also limit our attention to HHS data sets collected after 2005.

[7] Many of the nationally-representative multi-topic household surveys are known as the Living Standard Measurement Surveys (LSMS), but not all.

## 2.1 The World Bank Enterprise Surveys

The Enterprise Analysis Unit of the World Bank Group conducts three different types of surveys on enterprises around the world. The first is a standard establishment-level survey, known as the Enterprise Survey (henceforth, WBES Regular). The sampling protocol for the WBES Regular targets the formal private sector, and explicitly includes businesses with five or more employees. The sampling unit is the establishment — a business entity associated with a physical location with its own set of financial statements, including a balance sheet and income statement. The sampling protocols specifically stratify on industry sector and geographic location, however, they do not mention or discuss gender.

The second type is a survey of microenterprises, known as the Micro Enterprise Survey (henceforth, WBES Micro), which targets registered establishments with less than five employees. Sampling techniques and questionnaires are the same as the WBES Regular. For the WBES Micro, the regions covered are selected based on the number of establishments, contribution to employment, and value-added. In most cases, these regions are metropolitan areas and reflect the largest centers of economic activity in a country. Similar to the WBES Regular, enterprises covered under WBES Micro have a physical location and the sampling protocols stratify on industry sector and geographic location but do not explicitly mention or discuss gender.

The third type is the Informal Sector Enterprise Survey (henceforth, WBES Informal), which aims to provide information on a sample of informal private sector enterprises in selected urban centers within a country. WBES Informal has employed two different sampling methodologies over time that largely focus on visible businesses in dense business locations. See Appendix section C.1 for a detailed discussion. In either methodology, there is no mention of going to households and surveying informal businesses operating inside. Similar to the other WBES sampling protocols, there is no explicit gender lens.

**Key Variables** From all three types of WBES, we construct an indicator for the enterprise having at least one female owner[8] and the female- to male-owned enterprise sales ratio for each data set

[8]This is the most commonly collected indicator, though the WBES has several indicators related to women's participation in enterprise ownership. Other gender ownership variables collected include the percentage of the enterprise owned by women and if any of the *principal* or largest owners are female. Where the most commonly collected indicator is not available, we use these other indicators when possible to fill in if a business has at least one female owner.

using total annual enterprise sales.[9] Appendix sections C.2 and C.3 discuss the construction of sales ratio and addendum of sales in detail.

## 2.2 Household Surveys

We use multi-topic household surveys (HHS) that contain a module on non-farm enterprises.[10] The typical sampling strategy is a two-stage probability sample. First, areas from census-based sampling units (such as census tracts or enumeration districts) are chosen, and then dwellings from a list of all households within that sampling unit are chosen. These surveys are usually representative of the country as a whole, with large enough and stratified samples to allow consideration of certain subgroups, such as rural vs. urban, or a few major agro-climatic zones.

**Key Variables** Using the non-agricultural enterprises rostered in modules of HHS, we create a representative sample of enterprises found via households in a country in that given year.[11] For each enterprise in the roster, respondents list the business owners. We use the owner list to determine if there is at least one female owner (akin to the definition used in the WBES data) and the enterprise sales data to estimate the female- to male-owned enterprise sales ratio for each data set. We also create indicators of whether a business is formal (i.e., licensed or registered with the government, pays taxes, or registered with a tax collecting agency), located in a rural or urban area, and has a physical structure (i.e., operates business activities) outside of the home.

## 2.3 Other Data

**Ethiopia Large and Medium Manufacturing Industry Survey (LMMIS)** The Central Statistical Agency of Ethiopia has been conducting the LMMIS on an annual basis since 1996. The sample frame for the survey includes manufacturing enterprises that use power-driven machinery and engage 10 or more employees. It covers both public and private enterprises in all regions of the

[9] All sales information is converted into US dollars and adjusted according to the respective PPP exchange rate for 2020 and winsorized at the top 5% level. 18 data sets are excluded from sales-related analyses due to unusually high values and uncertainty about the required exchange rate adjustments.

[10] All the publicly available HHS data sets for Sub-Saharan Africa have a survey module on non-agricultural businesses owned by household members over the past 12 months, except Tanzania 2008 and 2010. We get the relevant non-farm business activities information from the 'Self-employment' section for those two data sets.

[11] It is important to note that the chance of the household-based sampling strategy producing a computationally meaningful sample of medium and large enterprise owners is low.

country. We use the 2006 and 2015 LMMIS data to calculate female ownership representation in Ethiopian manufacturing enterprises.

**Integrated Business Establishment Survey (IBES)** The Ghana Statistical Service conducted the IBES in 2014. The census identified 638,000 establishments across all sectors that had a physical structure and any household-based enterprise with a sign indicating its presence within a household. It excludes mobile businesses (hawkers), traders operating in temporary spaces, and household-based enterprises without visible signage. We use the 2014 Ghana economic census data to calculate the female ownership representation in non-farm enterprises in Ghana.

## 3 Female Ownership Representation in the WBES and HHS

Figure A1 provides estimates of the share of non-agricultural enterprises with at least one female owner from each data source across all countries and years surveyed. It depicts clear, systematic, and consistent differences in the female ownership estimates between the WBES surveys and the HHS data. An estimated 27.9% of enterprises in the WBES Regular have at least one female owner, while the estimated rate is 57.20%, 29.3 p.p higher, using the HHS data. The less frequently collected WBES Micro and Informal surveys also have visibly fewer women-owned enterprises compared to the HHS, with an estimated 35.10% and 37.8% of enterprises with at least one female owner in WBES Micro and Informal, respectively.[12] Ghana is the only Sub-Saharan African country that has a WBES Regular, WBES Informal, and an HHS all from the same year (2013). Rates of female ownership from these surveys are as follows: WBES Regular (29%), WBES Informal(63%), and HHS (70%).

We test for the statistical significance of the differences across these survey sources using the following specification:

$$Y_{ijs} = \beta_0 + \beta_1 WBESReg_{ijs} + \beta_2 WBESMicro_{ijs} + \beta_3 WBESInformal_{ijs} + \alpha_i + \gamma_j + \epsilon_{ijs} \quad (1)$$

where $Y_{ijs}$ is the estimated share of women-owned enterprises in country $i$ during survey year $j$ from data source $s$; $WBESReg_{ijs}$, $WBESMicro_{ijs}$, and $WBESInformal_{ijs}$ are binary indicators for the

[12] All observations are weighted using their respective survey source sampling weights. We also estimate the rate of female ownership without sampling weights, and our interpretation of findings does not change. Appendix Figure B1 presents the estimates.

data source of the estimates, with HHS as the reference group; $\alpha_i$ and $\gamma_j$ are fixed effects for the estimates' source country and year, respectively. Observations are at the country-year-data source level. All observations are weighted using their respective data source sampling weights.

Results are reported in Table 1. We find that female ownership representation in the HHS is significantly higher relative to all WBES sources and the differences are statistically significant. The inclusion of year and country fixed effects does not meaningfully alter the point estimates or statistical significance, indicating that these differences in female ownership rates are not being driven by the countries or years sampled. [13]

This large difference in women-owned enterprise representation relative to the HHS is not limited to the WBES. Appendix Figure A3 depicts results from a complementary analysis comparing the Ethiopia 2006 and 2015 WBES Regular data with the LMMIS data, and the Ghana 2013 WEBS Regular, WBES Informal, and HHS data with the IBES data. We find that the WBES Regular sampling protocol captures a similar percentage of women-owned enterprises in the manufacturing industry as the Ethiopia Manufacturing census (LMMIS) which focuses on formal businesses with 10 or more employees. Although we find that the Ghana 2014 Economic Census data yield considerably higher female ownership rates (56.65%) than the WBES Regular (29.63%), the census rate is still considerably and significantly lower than that of the HHS (71.22%). We posit this is likely due to the sampling criteria of the census that excludes businesses with no permanent physical structure and household-based enterprises without visible signage.

This comparison of women-owned enterprise representation in the Ethiopia LMMIS and Ghana IBES data to the WBES surveys and HHS suggests that the WBES Regular does a good job of representing the intended gender-blind sampling target (as shown in the comparison with the Ethiopia LMMIS). However, even the most comprehensive enterprise data of which we are aware (the Ghana IBES) has sampling criteria for which the average female entrepreneur is relatively less likely to meet than the average male entrepreneur. To the best of our knowledge, a comprehensive census of enterprise owners (both visible and "invisible", operating entirely within a household or as a mobile business) does not exist.

[13] These findings are also robust to alternative measures of female ownership (i.e., female manager or decision maker), as shown in Appendix Figure A2 and Table A2 as well as Appendix Figure B2 (unweighted). See Appendix section C.4 for the construction of the female manager indicator)

# 4 Gender Gap in Enterprise Performance in the WBES and HHS

Next, we examine how the sales ratio changes by data sources and test for the statistical significance of the differences by estimating equation 1, where $Y_{ijs}$ is now the estimated female- to male-owned enterprise sales ratio in country $i$ during survey year $j$ from data source $s$. All observations are still weighted using their respective data source sampling weights. [14]

Table 2 reports our findings, and Figure A4 complements this table with a visual representation. We find that female- and male-owned enterprises report almost equal sales in the WBES Regular (0.914) survey and it is statistically significantly different from the HHS average of 0.591.[15] The average female- to male-owned enterprise sales ratio estimate for WBES Micro and WBES Informal are 0.826 and 0.666, respectively. However, once we control for the source country and year-fixed effects, these estimates are not significantly different from the HHS estimate due to a high standard error.[16] [17]

# 5 What is Driving the Differences Across Data Sources?

To empirically explore the potential driving factors behind the different rates of female ownership across data sources, we compare female ownership representation within the HHS data sets across groups of enterprises that meet the WBES sampling protocol criteria compared to enterprises that do not. Specifically, we create two indicator variables 'Like WBES Micro' and 'Like WBES Informal' using the information on the total number of paid employees within an enterprise, formality status, rural or urban location, and physical location of the business (i.e., whether the business activities are operated from inside or outside of household). Registered enterprises that have less than 5

[14] We also estimate the female- to male-owned enterprise sales ratio without sampling weights, and our interpretation of findings does not change. Appendix Figure B3 presents the estimates.

[15] This echoes findings by (Bardasi, Sabarwal, and Terrell, 2011) that estimate little gender gaps in business performance using the WBES Regular.

[16] Appendix Table A1 shows that the WBES Informal surveys are conducted much less frequently. We also find a large number of missing sales information in these data sets compared to other survey sources: 54.37% of the sales data is missing.

[17] We do not look into the variation in sales ratio across data sources using the female manager or decision-maker variable, as there is a large amount of missing information. 44.87% of the WBES Regular has either the 'total annual sales' or 'female manager' information missing. The rate is 57.07%, 80.10%, and 32.01% for WBES Micro, WBES Informal, and HHS, respectively.

paid employees, operate business activities outside of household, and are located in the urban area are categorized as 'Like WBES Micro', whereas non-registered enterprises that operate business activities outside of household, and are located in the urban area are categorized as 'Like WBES Informal'.[18]

We test for the statistical significance of the differences across these HHS groups using the following specification:

$$Y_{eij} = \beta_0 + \beta_1 LikeMicro_{eij} + \beta_2 LikeInformal_{eij} + \beta_3 RuralInside_{eij} + \beta_4 RuralOutside_{eij} + \alpha_{ij} + \epsilon_{eij} \quad (2)$$

where $Y_{eij}$ is an indicator for female ownership of enterprise $e$ in country $i$ during survey year $j$; $LikeMicro_{eij}$ and $LikeInformal_{eij}$ are binary variables that indicate whether enterprise $e$ has characteristics akin to those in the WBES Micro and WBES Informal; $RuralInside_{eij}$ and $RuralOutside_{eij}$ refer to rural enterprises that operate business activities from inside and outside of the household, respectively; and $\alpha_{ij}$ is a data source fixed effect to account for country-year specific trends. The reference group consists of HHS enterprises without characteristics matching either the WBES Micro or WBES Informal. Standard errors are clustered at the data source level. All observations are weighted using a combined weight which is a multiplication of the survey source sampling weight and the ratio of the number of observations in each data source to the whole HHS data.

Table 3 columns 1 and 2 report our findings.[19] Enterprises within the HHS that satisfy the sampling criteria either of the Micro Enterprise or Informal Sector Enterprise Surveys are significantly less likely to have female owners compared to businesses that do not meet either sampling criteria. On average, only 33.6% of the 'Like WBES Micro' enterprises have at least one female owner, which is similar to the female ownership estimate of WBES Micro (32.3%) in Table 1 column 4, while the rate is 62.40% for the reference group. For 'Like WBES Informal', we estimate an average of 51.90% of enterprises with at least one female owner which is 10.5 p.p smaller than the reference group estimate. All these estimates are statistically different.

Given that WBES Micro and Informal surveys almost exclusively cover urban areas, one might suspect that the lower representation of female owners in these surveys might be explained by the location of the households in the HHS data (i.e., rural versus urban). However, our findings in Table 3 column 2 show that differences in the female ownership estimates across HHS enterprise

[18] Formal enterprises with five or more employees are rare in the HHS, without quantitatively meaningful numbers (only 292 out of 112,291 enterprises). We, therefore, exclude these enterprises from our analysis.

[19] Appendix Figure A5 presents a visual representation of the estimated rate of female ownership for these two HHS enterprise groups across all HHS countries and years surveyed.

groups are not explained by whether the business is in a rural or urban area but rather by whether the enterprise has a physical structure or not (i.e., operate business activities from inside or outside of household). There is no significant difference in female ownership estimates between rural and urban businesses that are operated from inside households. However, rural enterprises operated outside of the household have an average of 30.5 p.p lower female ownership estimate compared to urban enterprises operating inside of the household (the reference group). Combining the findings of columns 1 and 2, we conclude that the persistent gaps in relative rates of female ownership across the WBES Micro, WBES Informal, and HHS are likely primarily due to their differences in the other sampling target characteristics: for WBES Micro, it is driven by their focus on formal businesses and for WBES Informal, it is likely due to their focus on businesses with physical structure (i.e., operated outside of household).

We next examine how the difference in total annual sales of HHS enterprises across groups of enterprises that meet the sampling protocol of either the WBES Micro or WBES Informal varies by gender of the owner, using the following specification:

$$\begin{aligned} Y_{eij} =& \beta_0 + \beta_1 LikeMicro_{eij} + \beta_2 LikeInformal_{eij} + \beta_3 LikeMicro_{eij} * Female_{eij} + \\ & \beta_4 LikeInformal_{eij} * Female_{eij} + \beta_5 Female_{eij} + \alpha_{ij} + \epsilon_{eij} \end{aligned} \tag{3}$$

where $Y_{eij}$ is the total annual sales of enterprise $e$ in country $i$ during survey year $j$; $LikeMicro_{eij}$ and $LikeInformal_{eij}$ are binary variables that indicate whether enterprise $e$ has characteristics akin to those in the WBES Micro and WBES Informal, respectively; $Female_{eij}$ is an indicator of whether the enterprise has a female owner or not, and $\alpha_{ij}$ is a data source fixed effect to account for country-year specific trends. The reference group consists of male-owned enterprises that do not meet sampling frame characteristics of either the WBES Micro or the WBES Informal Survey – i.e., all male-owned enterprises located in the rural area and all male-owned urban enterprises that are operated inside the households. Standard errors are clustered at the data source level. All observations are weighted using a combined weight which is a multiplication of the survey source sampling weight and the ratio of the number of observations in each data source to the whole HHS data. Table 3 column 3 reports the findings.

We find that, on average, HHS businesses that are women-owned generate significantly less annual sales compared to male-owned enterprises in the reference group. However, women-owned businesses that comprise 'Like Micro' and 'Like Informal' have on average of 2,330.36 and 952.43 USD more annual sales compared to 'Like Micro' and 'Like Informal' male-owned businesses,

respectively. These differences imply that women-owned businesses with characteristics similar to the sampling protocol of either the WBES Micro or WBES Informal are more positively selected on performance relative to their respective male-owned enterprise groups. This pattern could be driven by underlying occupational choice fundamentals that more easily allow men to exit this specific type of employment for better labor market opportunities (Hardy, Litzow, McCasland, and Kagy, 2023).

## 6 Differing Policy Implications from Different Data Sources

We use the information on the self-reported challenges enterprises face from the WBES and HHS data sets and generate an estimate of their implied resource preferences from each data source. The objective of this analysis is to understand, from a policymaker's or a researcher's point of view, how each data source may imply that resources or researchers' attention should be allocated across potential problems that enterprises may be facing. We construct an implied resource preference index for 8 broad categories: *Infrastructure*, *Market Issues*, *Capital*, *Governance*, *Safety*, *Technology*, *Labor* and *Land*. A particular survey respondent's answers to business constraint questions are translated into an implied resource allocation preference across these 8 constraint categories, depending on the number and severity of each reported constraint. For example, if the survey asked about the three major business barriers the enterprise is facing and the respondent list one constraint categorized under *Market Issues* and two constraints categorized under *Governance*, then that respondent's implied resource preference index is 33.33% for *Market Issues*, 66.66% for *Governance*, and 0% for all other categories.[20]

We test for the statistical significance of the differences in implied resources preference for each of the 8 broad categories across data sources using equation 1, $Y_{ijs}$ is the estimated implied resource preference for a particular constraint category in country $i$ during survey year $j$ from data source $s$. Observations are at the country-year-data source level. All observations are weighted using their

[20] Appendix Table A3 summarizes the list of constraints listed across different data sources and the categorization of the 8 broad categories. Appendix C.6 explains the details of the standardized implied resource preference index construction. Due to the considerable variations in the survey structure (i.e., questions asked on business constraints and options listed across data sources), there is no definitive way in which one can construct the implied resource preference index. As a robustness check, we discuss an alternative way of constructing the resource preference in Appendix C.7. Appendix Table A4 reports the estimated resource preference using this alternative index and the findings are similar to our findings in Table 4 described below.

respective survey source sampling weights. Panel A of Table 4 presents the results. We find that the WBES point toward a different set of enterprise-focused policy and research priorities than those highlighted by the HHS. Column 1 shows that HHS businesses, on average, would prefer to have 38.2% of total resources allocated towards resolving market access and information-related problems. The rate is statistically significantly lower in the WBES Regular, WBES Micro, and WBES Informal – 7.6%, 8.4%, and 4.6%, respectively. On the other hand, Column 4 shows that WBES Regular and WBES Micro enterprises would, on average, prefer to have 29.9% and 27.9% of total resources allocated towards resolving governance-related issues, respectively, whereas HHS enterprises would like to have an allocation of only 4.4% towards that category. In summary, the HHS data implies that market access, infrastructure, and capital are the key constraints faced by enterprises. In contrast, WBES data suggests that governance and safety issues are the key business barriers in SSA.[21]

The findings of Panel A can be driven by either the differences in sampling protocols documented and discussed above or by the differences in constraint questions asked about in each survey – i.e., the survey structure. For example, constraints that comprise the 'Market' category are mostly available in the HHS – the only option listed in WBES surveys is "Informal Firm Practice". On the other hand, none of the HHS data except Uganda include 'Land' as a constraint. To empirically investigate how much of the variations in the implied resource preferences can be explained by this variation in survey structures, we create a randomized set of responses to the constraint-related questions available in each data source and repeat the Panel A analysis. Random generation of responses effectively removes the part of the answer driven by the actual respondent's preferences and leaves only the answer propensity implied by the survey structure itself. A significant correlation between data source indicators and constraint categories would indicate that survey structure plays an important role in determining which constraints appear to be more prevalent. A significant difference between the coefficient estimates generated from true response data and those estimates generated from random response data implies that there remains explanatory variation in Panel A that is due to the sampling frame rather than survey construction.

Panel B of Table 4 reports the findings. Panel C of Table 4 reports the P-value of the difference in point estimates between Panel A and B. We find that differences in survey construction is an important factor in determining which constraint category gets more implied resource preference

[21] Note that this echoes the findings by (Eifert et al., 2008) that claim poor governance and infrastructure to be the major barriers to competitiveness for manufacturing enterprises in Sub-Saharan African countries using the WBES Regular.

– many of the data source indicators in Panel B are statistically significantly different from zero for the different constraint categories. We also find that for the 'Market' and 'Safety' categories, Panel A and Panel B estimates are significantly different, indicating that the differences in implied resource preference are driven both by sampling protocols and survey structure. Overall, our findings demonstrate the importance of considering both survey and sampling protocol design in the enterprise-focused policy implications derived from these data sources.

# 7 Conclusion

In order for evidence to effectively drive equity-focused policies and research, it is necessary to have data that accurately represents our society. This paper highlights the importance of considering representation in sampling protocols of common data sources. Examining the case of women-owned enterprise representation, we detect large variations in the share of women-owned businesses, diagnostics of gender gaps in business performance, and the implications implied for policy and research priorities across commonly available data sources.

We posit that the issues of representation exposited through the specific enterprise gender data gap documented in this paper are not limited to enterprise data or gender. For example, in the economics profession, it is quite common to "drop the women" due to the costs of inclusion, whether it be data collection costs or modeling complications from the dynamic nature of their labor market involvement. For example, economic history commonly focuses on men due to the substantial challenge in linking historical records of women who changed their last name at the time of marriage and the high upfront costs of digitizing marriage records (Bailey et al., 2017; Jácome et al., 2021). Similarly, medical research overwhelmingly excludes pregnant women, sometimes extended to pregnable, and once-pregnable women (Merton, 1993; Michelman and Msall, 2021; Murthy et al., 2004). Such underrepresentation of women has been shown to lead to biased economic estimates and distort resulting knowledge and policy learning (Merton, 1993; Michelman and Msall, 2021). Beyond gender representation, other research has shown how political-economic factors that drive sample selection issues have also led to unbalanced representation, leading to national resource distortion and ineffective policy choice.[22]

Our specific empirical findings make clear that existing data on enterprises and their owners

[22] Wang and Yang (2021) in the context of China; Alsan, Durvasula, Gupta, Schwartzstein, and Williams (2022) and Allcott (2015) in the context of the U.S.

are sampled in such a way that it is less informative about the experiences and needs of the average female enterprise owner, relative to that of the average male owner. It is essential to consider these gaps when using existing enterprise data to study gender and discuss possibilities for mitigating such gaps in future data collection. More broadly, as a profession, it is imperative to be aware that these seemingly benign sampling decisions in data collection have large implications for representation bias in our understanding of economic behavior and the design of equitable and inclusive policy, whether we are the collectors of such data or it's secondary users.

## References


**Abor, Joshua, and Peter Quartey.** 2010. "Issues in SME development in Ghana and South Africa." *International research journal of finance and economics* 39 (6): 215–228.

**Allcott, Hunt.** 2015. "Site selection bias in program evaluation." *The Quarterly Journal of Economics* 130 (3): 1117–1165.

**Alsan, Marcella, Maya Durvasula, Harsh Gupta, Joshua Schwartzstein, and Heidi L Williams.** 2022. "Representation and Extrapolation: Evidence from Clinical Trials."Technical report, National Bureau of Economic Research.

**Bailey, Martha Jane, Sarah Anderson, and Catherine Massey.** 2017. "Life-m: The longitudinal, intergenerational family electronic micro-database." In *PAA 2017 Annual Meeting*, PAA.

**Bardasi, Elena, Shwetlena Sabarwal, and Katherine Terrell.** 2011. "How do female entrepreneurs perform? Evidence from three developing regions." *Small Business Economics* 37 (4): 417–441.

**Delecourt, Solène, and Anne Fitzpatrick.** 2021. "Childcare matters: Female business owners and the baby-profit gap." *Management Science* 67 (7): 4455–4474.

**Delecourt, Solène, and Odyssia Ng.** 2021. "Does gender matter for small business performance? Experimental evidence from India." unpublished.

**Eifert, Benn, Alan Gelb, and Vijaya Ramachandran.** 2008. "The cost of doing business in Africa: Evidence from enterprise survey data." *World development* 36 (9): 1531–1546.

**Enterprise Analysis Unit, World Bank Group.** 2021. "Enterprise Surveys Manual and Guide."Technical report, World Bank.

**Fang, Sheng, Chorching Goh, Mark Roberts, L Colin Xu, and Albert Zeufack.** 2020. "Female business leaders, business and cultural environment, and productivity around the world." *World Bank Policy Research Working Paper* . (9275): .

**Gindling, TH, and David Newhouse.** 2014. "Self-employment in the developing world." *World Development* 56 313–331.

**Hallward-Driemeier, Mary, and Lant Pritchett.** 2015. "How business is done in the developing world: Deals versus rules." *Journal of economic perspectives* 29 (3): 121–40.

**Hardy, Morgan, and Gisella Kagy.** 2018. "Mind the (profit) gap: why are female enterprise owners earning less than men?" In *AEA Papers and Proceedings*, Volume 108. 252–55.

**Hardy, Morgan, and Gisella Kagy.** 2020. "It's Getting Crowded in Here: Experimental Evidence of Demand Constraints in the Gender Profit Gap." *The Economic Journal* 130 (631): 2272–2290.

**Hardy, Morgan, Erin Litzow, Jamie McCasland, and Gisella Kagy.** 2023. "Gender Differences in Informal Labor-Market Resilience." *The World Bank Economic Review* 37 (1): 112–126.

**Henrich, Joseph, Steven J Heine, and Ara Norenzayan.** 2010. "The weirdest people in the world?" *Behavioral and brain sciences* 33 (2-3): 61–83.

**Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu.** 2021. "Mobility for all: Representative intergenerational mobility estimates over the 20th century."Technical report, National Bureau of Economic Research.

**Kerr, Andrew, and Bruce McDougall.** 2020. "What is a firm census in a developing country? An answer from Ghana."Technical report, Centre for the Study of African Economies, University of Oxford.

**Merton, Vanessa.** 1993. "The exclusion of pregnant, pregnable, and once-pregnable people (aka women) from biomedical research." *American Journal of Law & Medicine* 19 (4): 369–451.

**Michelman, V., and L. Msall.** 2021. "Sex, Drugs, and R&D: Missing Innovation from Regulating Female Enrollment in Clinical Trials." unpublished.

**Murthy, Vivek H, Harlan M Krumholz, and Cary P Gross.** 2004. "Participation in cancer clinical trials: race-, sex-, and age-based disparities." *Jama* 291 (22): 2720–2726.

**Nix, Emily, Elisa Gamberoni, and Rachel Heath.** 2015. "Bridging the Gender Gap: Identifying What Is Holding Self-employed Women Back in Ghana, Rwanda, Tanzania, and the Republic of Congo." *The World Bank Economic Review* 30 (3): 501–521.

**Riley, Emma.** 2020. "Resisting social pressure in the household using mobile money: Experimental evidence on microenterprise investment in Uganda." *University of Oxford* 25.

**UNDP, United Nations Development Programme.** 2015. *The Sustainable Development Goals (SDGs)*. UNDP.

**Wang, Shaoda, and David Y Yang.** 2021. "Policy Experimentation in China: The Political Economy of Policy Learning."Technical report, National Bureau of Economic Research.

**World Bank Group.** 2019. *Profiting from Parity: Unlocking the Potential of Women's Business in Africa*. World Bank.

## Exhibits

Table 1: Female Ownership by Survey Type

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| WBES Regular | -0.293*** | -0.299*** | -0.319*** | -0.332*** |
| | (0.019) | (0.023) | (0.019) | (0.023) |
| WBES Micro | -0.221*** | -0.226*** | -0.256*** | -0.275*** |
| | (0.033) | (0.036) | (0.032) | (0.036) |
| WBES Informal | -0.194*** | -0.190*** | -0.253*** | -0.252*** |
| | (0.034) | (0.036) | (0.035) | (0.038) |
| Year dummies | No | Yes | No | Yes |
| Country dummies | No | No | Yes | Yes |
| HHS Mean | 0.572 | 0.572 | 0.572 | 0.572 |
| P-Value (Reg=Micro) | 0.029 | 0.037 | 0.025 | 0.043 |
| P-Value (Reg=Inf) | 0.004 | 0.002 | 0.040 | 0.022 |
| P-Value (Micro=Inf) | 0.540 | 0.418 | 0.927 | 0.547 |
| Observations | 168 | 168 | 168 | 168 |
| R-sqr | 0.491 | 0.517 | 0.775 | 0.798 |

Note: This table reports the regression of female ownership on survey type. HHS is the base category and each coefficient reflects the difference in means of female ownership between HHS and respective survey types. Column 1 shows the simple OLS regression with no fixed effects. Column 2 includes year fixed effects, column 3 includes country fixed effects, and column 4 includes country fixed effects and year fixed effects. All observations are weighted using their respective survey source sampling weights. Robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 2: Female to Male Sales Ratio by Survey Type

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| WBES Regular | 0.324*** | 0.291*** | 0.313*** | 0.283** |
| | (0.055) | (0.079) | (0.098) | (0.124) |
| WBES Micro | 0.235*** | 0.312** | 0.139 | 0.246 |
| | (0.090) | (0.136) | (0.114) | (0.157) |
| WBES Informal | 0.076 | 0.138 | 0.038 | 0.070 |
| | (0.059) | (0.094) | (0.135) | (0.166) |
| Year dummies | No | Yes | No | Yes |
| Country dummies | No | No | Yes | Yes |
| HHS Mean | 0.591 | 0.591 | 0.591 | 0.591 |
| P-Value (Reg=Micro) | 0.386 | 0.868 | 0.096 | 0.752 |
| P-Value (Reg=Inf) | 0.001 | 0.138 | 0.016 | 0.102 |
| P-Value (Micro=Inf) | 0.125 | 0.259 | 0.458 | 0.263 |
| Observations | 149 | 149 | 149 | 149 |
| R-sqr | 0.124 | 0.271 | 0.407 | 0.537 |

Note: This table reports regression of ratio of female- to male-owned enterprises sales on survey type. HHS is the base category and the coefficients reflect the difference in means of sales ratio between HHS and respective survey types. Column 1 shows the simple OLS regression with no fixed effects. Column 2 includes year fixed effects, column 3 includes country fixed effects, and column 4 includes country and year fixed effects. 18 data sources are excluded from this analysis due to unusually high sales value and uncertainty about the required exchange rate adjustments. Appendix C.3 discusses this in detail. All observations are weighted using their respective survey source sampling weights. Robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table 3: Female Ownership and Sales within the HHS

| | (1) Female Ownership | (2) Female Ownership | (3) Total Sales |
|---|---|---|---|
| Like WBES Micro | -0.288** | -0.417*** | 7869.834*** |
| | (0.100) | (0.093) | (2092.257) |
| Like WBES Informal | -0.105** | -0.232*** | 2050.385** |
| | (0.041) | (0.034) | (673.048) |
| Rural, Operated Inside | | -0.020 | |
| | | (0.018) | |
| Rural, Operated Outside | | -0.305*** | |
| | | (0.015) | |
| Like WBES Micro*Female-Owned | | | 2330.359* |
| | | | (1261.765) |
| Like WBES Informal*Female-Owned | | | 952.431** |
| | | | (367.396) |
| Female-Owned | | | -2423.084*** |
| | | | (298.122) |
| Reference Group Mean | 0.624 | 0.782 | 6503.724 |
| Observations | 87515 | 87515 | 87515 |
| R-sqr | 0.026 | 0.085 | 0.433 |

Note: This table shows the regression of female ownership and total annual sales in dollars (PPP adjusted) on different indicators. The first three columns use HHS data. The fourth column uses the Ghana IBES census data. The first three regressions control for data source fixed effects. The last column uses data for one country and one year thus no fixed effects are included. For the first three columns, enterprises that satisfy the criteria of WBES Regular are excluded from the analysis as they constitute only 0.29% of total observations. The reference group for the first column consists of all enterprises located in the rural area and all urban enterprises operated inside the households. For the second column, the reference group consists of all urban enterprises operated inside the households. The third column reference group includes all male-owned enterprises located in the rural area and all male-owned urban enterprises operated inside the households. The fourth column reference group is all unregistered firms. Observations are weighted using a combination of survey weight and the number of observations on each country-year survey compared to the total number of observations in the HHS. Standard errors are clustered at the data source level and reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01

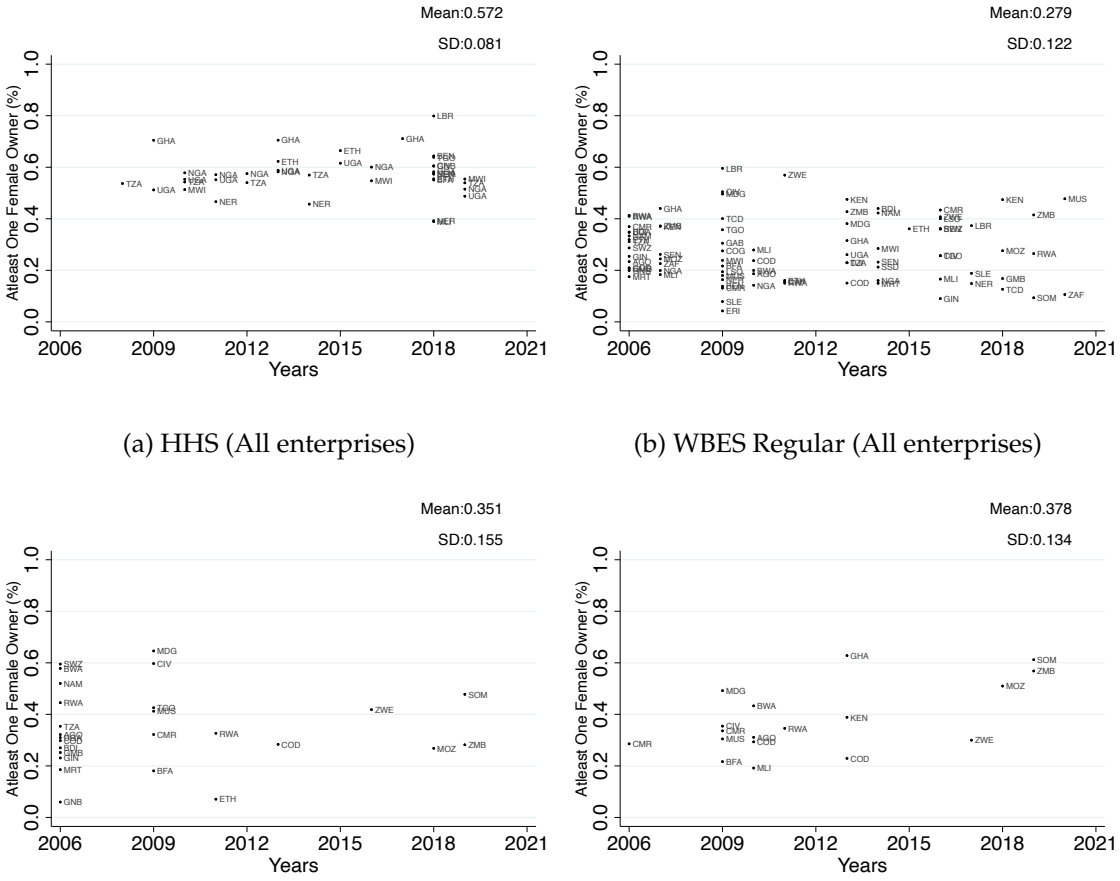
Table 4: Implied Resource Priority by Survey Type

| | (1) Market | (2) Infra. | (3) Capital | (4) Gov. | (5) Safety | (6) Tech. | (7) Labor | (8) Land |
|---|---|---|---|---|---|---|---|---|
| **Panel A: Original Constraint Responses** | | | | | | | | |
| WBES Regular | -0.306*** | -0.095 | -0.051 | 0.255*** | 0.164*** | -0.018*** | 0.024*** | 0.057*** |
| | (0.048) | (0.073) | (0.034) | (0.018) | (0.025) | (0.004) | (0.008) | (0.008) |
| WBES Micro | -0.298*** | -0.105 | -0.009 | 0.235*** | 0.150*** | -0.017*** | 0.007 | 0.067*** |
| | (0.048) | (0.074) | (0.038) | (0.022) | (0.026) | (0.004) | (0.009) | (0.009) |
| WBES Informal | -0.336*** | -0.129 | 0.208*** | 0.003 | 0.186*** | -0.015*** | -0.008 | 0.119*** |
| | (0.054) | (0.083) | (0.051) | (0.028) | (0.041) | (0.005) | (0.012) | (0.014) |
| HHS Mean | 0.382 | 0.312 | 0.171 | 0.044 | 0.029 | 0.016 | 0.013 | 0.000 |
| R-sqr | 0.849 | 0.641 | 0.720 | 0.907 | 0.788 | 0.681 | 0.665 | 0.809 |
| **Panel B: Randomly Generated Constraint Responses** | | | | | | | | |
| WBES Regular | -0.089*** | -0.163*** | -0.048** | 0.266*** | 0.107*** | -0.080*** | 0.014 | 0.056*** |
| | (0.009) | (0.025) | (0.021) | (0.018) | (0.019) | (0.009) | (0.016) | (0.006) |
| WBES Micro | -0.087*** | -0.168*** | -0.049** | 0.264*** | 0.098*** | -0.079*** | 0.026 | 0.057*** |
| | (0.009) | (0.027) | (0.020) | (0.019) | (0.021) | (0.009) | (0.017) | (0.007) |
| WBES Informal | -0.113*** | -0.140*** | 0.031 | -0.016 | 0.207*** | -0.076*** | 0.049 | 0.120*** |
| | (0.015) | (0.036) | (0.024) | (0.033) | (0.032) | (0.009) | (0.033) | (0.012) |
| HHS Mean | 0.157 | 0.320 | 0.142 | 0.101 | 0.082 | 0.087 | 0.046 | 0.007 |
| R-sqr | 0.843 | 0.739 | 0.725 | 0.924 | 0.743 | 0.890 | 0.476 | 0.828 |
| Observations (Panel A and B) | 146 | 146 | 146 | 146 | 146 | 146 | 146 | 146 |
| **Panel C: P-Value of the Difference in Point Estimates between Panel A and Panel B** | | | | | | | | |
| WBES Regular | 0.0000 | 0.2180 | 0.8511 | 0.4184 | 0.0002 | 0.0000 | 0.2520 | 0.8638 |
| WBES Micro | 0.0000 | 0.2550 | 0.0984 | 0.1028 | 0.0020 | 0.0000 | 0.0765 | 0.2049 |
| WBES Informal | 0.0000 | 0.8429 | 0.0000 | 0.3089 | 0.3169 | 0.0000 | 0.0100 | 0.9132 |

Note: This table shows the regression of implied resource priorities for different enterprise constraints on survey type. Panel A shows the implied intensity of constraints constructed from original responses whereas Panel B shows the findings from randomly generated constraint responses. Panel C reports the P-value of the difference in point estimates between Panel A and B. Appendix Section C.6 explains the details of the variable construction. All observations are weighted using their respective survey source sampling weights. HHS is the reference group and each coefficient reflects the difference in means of implied resource preference for a particular constraint between HHS and the respective survey source type. All regressions include fixed effects for the estimates source country and year, respectively to account for time-invariant country characteristics and yearly trends. Robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01

# A Appendix

Figure A1: Female Ownership Representation by Data Source, Country and Survey Year




(a) HHS (All enterprises)

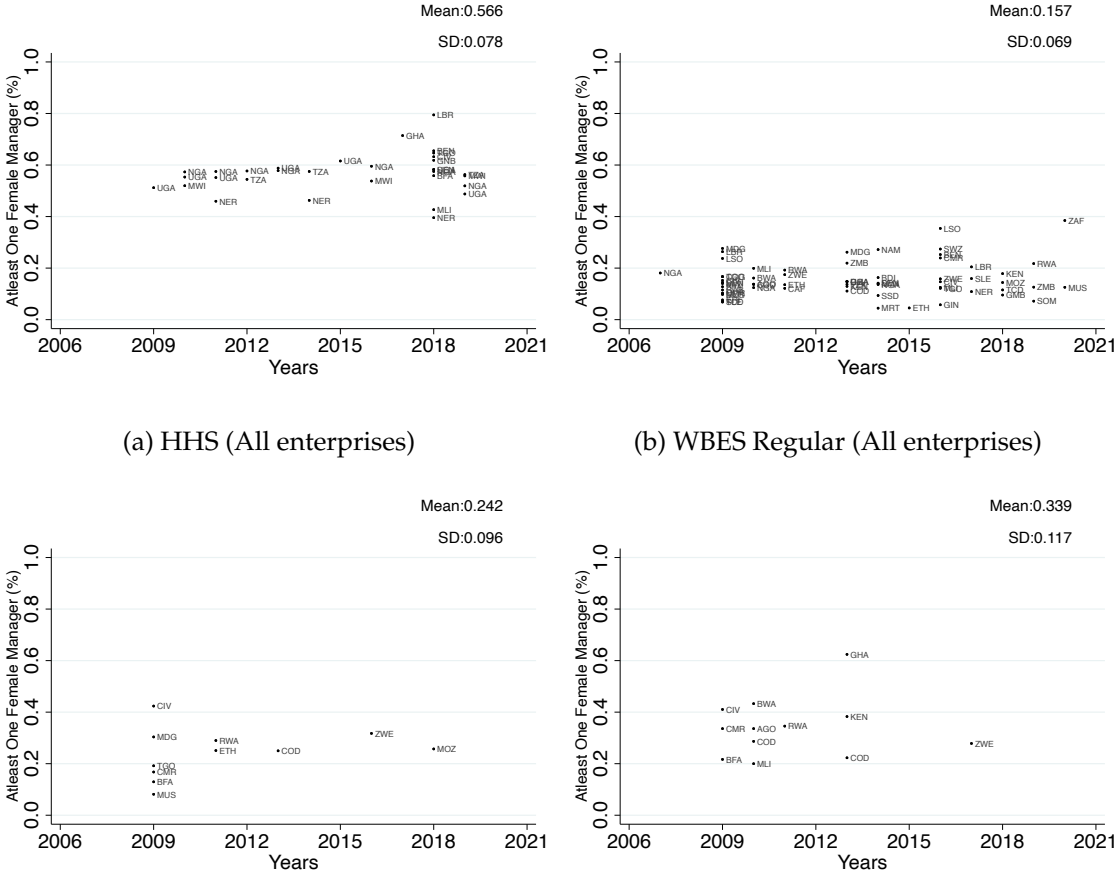
(b) WBES Regular (All enterprises)

(c) WBES Micro Survey

(d) WBES Informal Survey

Note: This figure shows the share of enterprises reporting at least one female owner by country and year of survey. Figure (a) shows data from 39 Multi-topic Household Surveys (HHS) covering 15 countries. Figure (b) shows data from 85 World Bank Enterprise Surveys (WBES) covering 43 countries. Figure (c) shows data from 26 Micro Enterprise Surveys (WBES Micro) covering 24 countries. Figure (d) shows data from 18 Informal Sector Enterprise Surveys (WBES Informal) covering 15 countries. All observations are weighted using their respective survey source sampling weights.

Figure A2: Female Manager/Decision Maker Representation by Data Source, Country and Survey Year




(a) HHS (All enterprises)

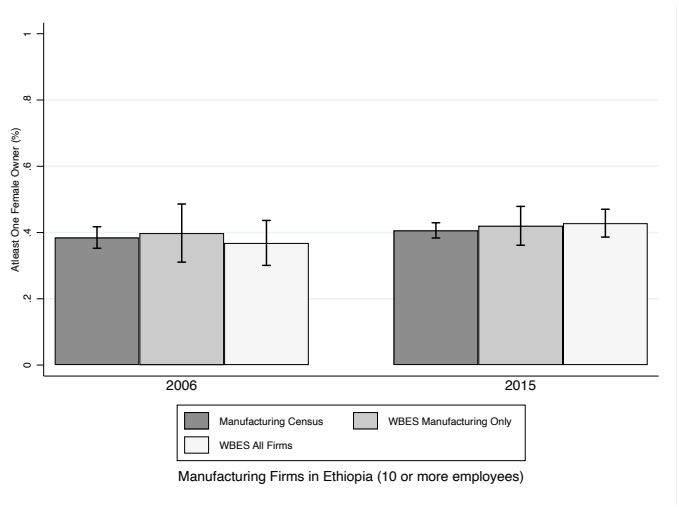
(b) WBES Regular (All enterprises)

(c) WBES Micro Surveys

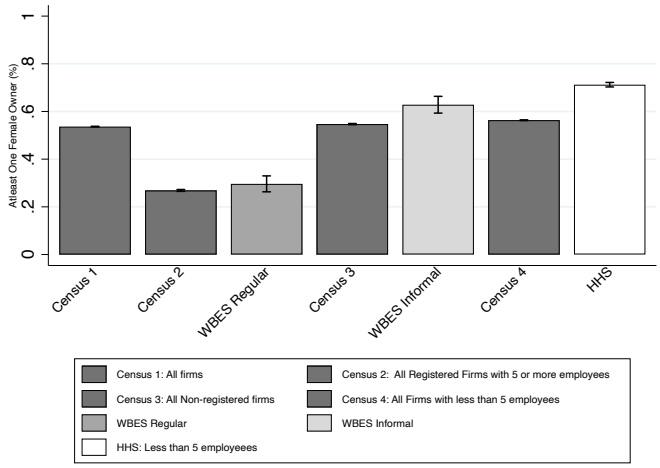
(d) WBES Informal Surveys

Note: This figure shows the share of enterprises reporting at least one female manager or decision maker by country and year of the survey. Figure (a) shows data from 38 Multi-topic Household Surveys (HHS) covering 15 countries. Figure (b) shows data from 85 World Bank Enterprise Surveys (WBES) covering 43 countries. Ethiopia is excluded from this analysis (Figure a) because there is no information on a business manager. Figures (c) and (d) show data from WBES Micro and Informal, respectively. Note that enterprises in WBES Micro and WBES Informal have a large number of missing values on enterprise manager gender (51.10% and 49.44%, respectively). All observations are weighted using their respective survey source sampling weights.

Figure A3: Female Ownership Representation In Ethiopia and Ghana



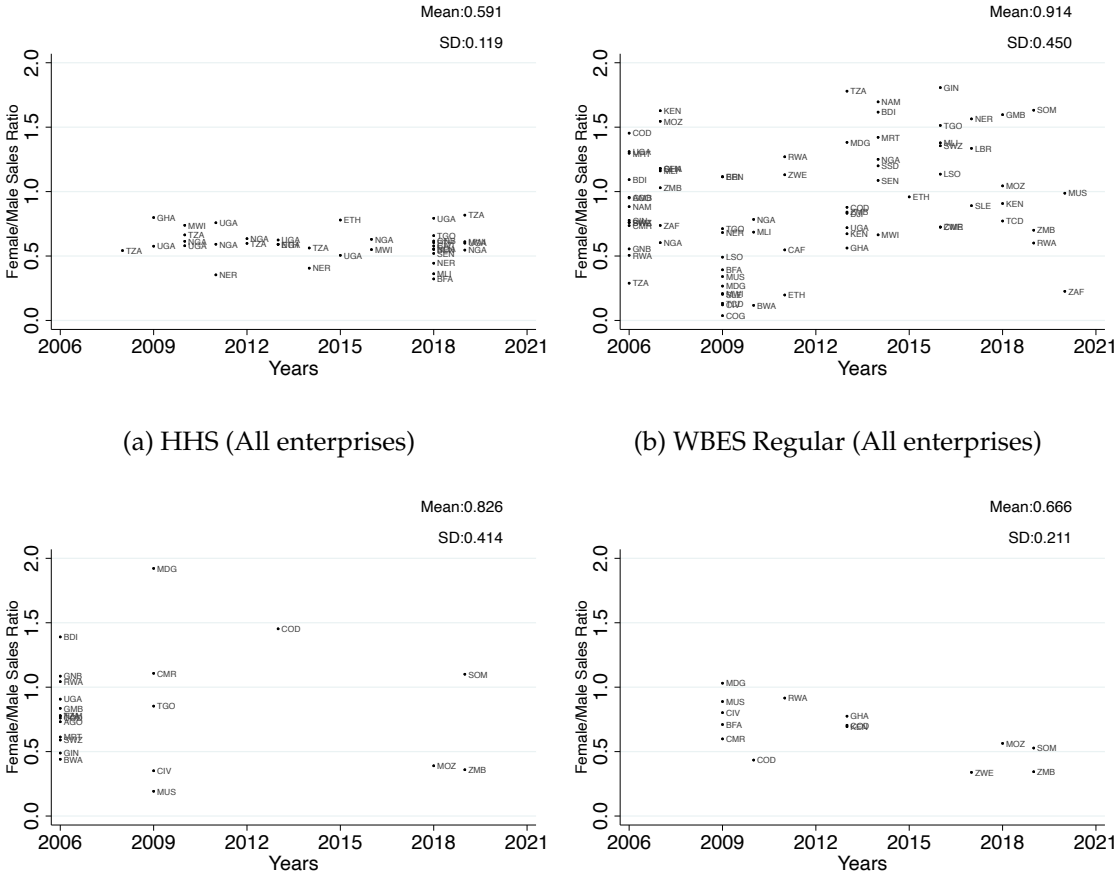

(a) Ethiopia Census




(b) Ghana Census

Note: This figure shows female ownership distribution for two case study countries, Ethiopia and Ghana. Figure (a) shows the rate of female ownership in formal private businesses by data source and survey year in Ethiopia – The Large and Medium Manufacturing Industry Survey (LMMIS) and the WBES Regular. There are two estimates of female ownership representation for WBES Regular– i) for manufacturing enterprises only, ii) for all enterprises. Enterprises with less than 10 employees are excluded from the analysis because the LMMIS report states that the survey only covers enterprises that engage 10 people or more and use power-driven machinery. Figure (b) shows the rate of female ownership in private non-farm businesses by data source in Ghana – an establishment census known as the Integrated Business Establishment Survey (IBES) 2014, WBES Regular 2013, WBES Informal survey 2013, and HHS 2013. Different subsets of the IBES data were considered with conditions akin to those in the sampling frame of the WBES Regular, WBES Informal, and HHS datasets. Appendix section C.5 describes how enterprise size is defined in different surveys.

Figure A4: Female to Male Sales Ratio by Data Source, Country and Survey Year




(a) HHS (All enterprises)

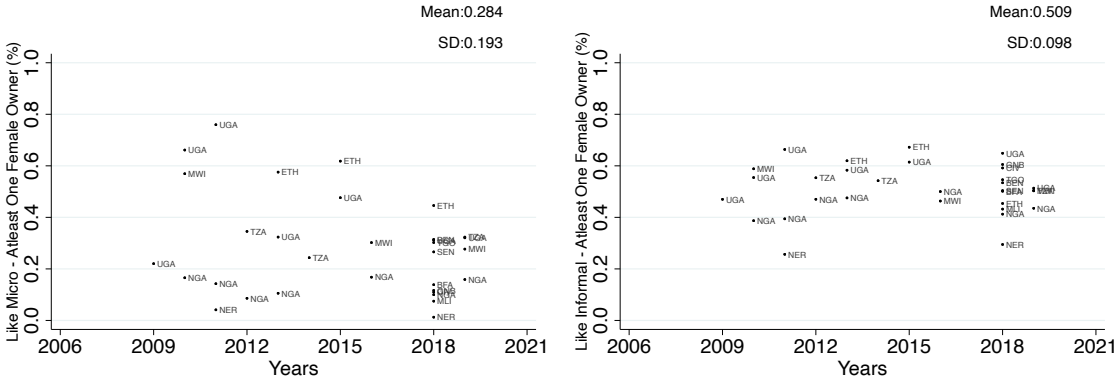
(b) WBES Regular (All enterprises)

(c) WBES Micro Survey

(d) WBES Informal Survey

Note: This figure shows the ratio of female- to male-owned enterprise sales by country and year of survey. Figure (a) shows data for all enterprises from the HHS survey. Figure (b) shows data for all enterprises from the WBES Regular. Figure (c) shows data for all enterprises from the WBES Micro survey. Figure (d) shows data for all enterprises from the WBES Informal survey. Sales are converted into US dollars, adjusted by the PPP exchange rate 2020, and winsorized at the top 5% level. 18 data sources are excluded from this analysis due to unusually high sales value and uncertainty about the required exchange rate adjustments. Appendix C.3 discusses this in detail. All observations are weighted using their respective survey source sampling weights.

Figure A5: Female Ownership Representation in HHS Sub-samples




(a) Analogue to WBES Micro

(b) Analogue to WBES Informal

Note: This figure shows the rate of female ownership across two groups of HHS enterprises – those who meet the sampling frame characteristics of either WBES Micro or WBES Informal – by country and year of the survey. Figure (a) shows data for HHS businesses that have less than 5 employees, operate outside a household, are registered with the government, and are located in an urban area, just like the enterprises considered for WBES Micro. Figure (b) shows data for businesses that operate outside a household, are not registered with the government and are located in an urban area, just like the enterprises considered for WBES Informal. All observations are weighted using their respective survey source sampling weights.

Table A1: List of Survey Years for Sub-Saharan African countries

| Country | WBES | | | HHS | Census |
|---|---|---|---|---|---|
| | Regular | Micro | Informal | | |
| Angola | 2006, 2010 | 2006 | 2010 | | |
| Benin | 2009, 2016 | | | 2018 | 1981 [c], 2008 [c] |
| Botswana | 2006, 2010 | 2006 | 2010 | | |
| Burkina Faso | 2009 | 2009 | 2009 | 2018 | |
| Burundi | 2006, 2014 | 2006 | | | |
| Cameroon | 2006, 2009, 2016 | 2009 | 2006, 2009 | | 2009[b] |
| CAR | 2011 | | | | |
| Chad | 2009, 2018 | | | | 2015[c] |
| DRC | 2006, 2010, 2013 | 2006, 2013 | 2010, 2013 | | |
| Republic of Congo | 2009 | | | | |
| Cote d'Ivoire | 2009, 2016 | 2009 | 2009 | 2018 | |
| Djibouti | 2013 | | | | |
| Eritrea | 2009 | | | | |
| Eswatini | 2006, 2016 | 2006 | | | 2011[c] |
| Ethiopia | 2006, 2011, 2015 | 2011 | | 2013, 2015, 2018 | 1996-2020 [i] |
| Gabon | 2009 | | | | |
| Gambia | 2006, 2018 | 2006 | | | |
| Ghana | 2007, 2013 | | 2013 | 2009, 2013, 2017 | 1962[b],1977[b], 1987 [b], 2003[b], 2014 [b] |
| Guinea | 2006, 2016 | 2006 | | | |
| Guinea-Bissau | 2006 | 2006 | | 2018 | |
| Kenya | 2007, 2013, 2018 | | 2013 | | 2012[b], 2017 [b] |
| Lesotho | 2009, 2016 | | | | 2012[b], 2015 [b] |
| Liberia | 2009, 2017 | | | 2018 | 2007 [c] |

[b] Business census - Covers all formal businesses

[c] Economic census - Covers all formal and informal businesses

[i] Industry census - Covers all formal manufacturing industries

Table A1 continued: List of Survey Years for Sub-Saharan African countries

| Country | WBES | | | HHS | Census |
|---|---|---|---|---|---|
| | Regular | Micro | Informal | | |
| Madagascar | 2009, 2013 | 2009 | 2009 | | |
| Malawi | 2009, 2014 | | | 2010, 2016, 2019 | |
| Mali | 2007, 2010, 2016 | | 2010 | 2018 | |
| Mauritania | 2006, 2014 | 2006 | | | |
| Mauritius | 2009, 2020 | 2009 | 2009 | | |
| Mozambique | 2007, 2018 | 2018 | 2018 | | |
| Namibia | 2006, 2014 | 2006 | | | |
| Niger | 2009, 2017 | | | 2011, 2014, 2018 | |
| Nigeria | 2007, 2010, 2014 | | | 2010, 2011, 2012, 2013, 2016, 2018, 2019 | |
| Rwanda | 2006, 2011, 2019 | 2006, 2011 | 2011 | | 2011 [c] , 2014 [c] , 2017 [c], 2020 [c] |
| Senegal | 2007, 2014 | | | 2018 | 2017 [c] |
| Sierra-Leone | 2009, 2017 | | | | 2005 [b] , 2016 [b] |
| Somlia | 2019 | 2019 | 2019 | | |
| South Africa | 2007, 2020 | | | | |
| South Sudan | 2014 | | | | |
| Tanzania | 2006, 2013 | 2006 | | 2008, 2010, 2012, 2014, 2019 | 2011-2012 [b] |
| Togo | 2009, 2016 | 2009 | | 2018 | |
| Uganda | 2006, 2013 | 2006 | | 2009, 2010, 2011, 2013, 2015, 2018, 2019 | 2001 [b] |
| Zambia | 2007, 2013, 2019 | 2019 | 2019 | | |
| Zimbabwe | 2011, 2016 | 2016 | 2017 | | |

[b] Business census - Covers all formal businesses

[c] Economic census - Covers all formal and informal businesses

[i] Industry census - Covers all formal manufacturing industries

Table A2: Female Manager/Decision Maker Representation by Survey Type

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| WBES Regular | -0.410*** | -0.414*** | -0.423*** | -0.412*** |
| | (0.016) | (0.016) | (0.018) | (0.022) |
| WBES Micro | -0.324*** | -0.321*** | -0.332*** | -0.309*** |
| | (0.031) | (0.030) | (0.032) | (0.035) |
| WBES Informal | -0.227*** | -0.225*** | -0.245*** | -0.226*** |
| | (0.036) | (0.036) | (0.031) | (0.036) |
| Year dummies | No | Yes | No | Yes |
| Country dummies | No | No | Yes | Yes |
| HHS Mean | 0.566 | 0.566 | 0.566 | 0.566 |
| Observations | 118 | 118 | 118 | 118 |
| R-sqr | 0.834 | 0.849 | 0.945 | 0.951 |

Note: This table reports the regression of female managers or decision makers on survey type. HHS is the base category and each coefficient reflects the difference in means of female managers or decision makers between HHS and respective survey types. Column 1 shows the simple OLS regression with no fixed effects. Column 2 controls for year fixed effects, column 3 controls for country fixed effects, and column 4 controls for country and year fixed effects. All observations are weighted using their respective survey source sampling weights. Robust standard errors are in parentheses.
* p<0.10, ** p<0.05, *** p<0.01

# Table A3: Constraint Broad Categories Construction

| | Market | Capital | Land | Labor | Infrastructure | Safety | Governance | Technology |
|---|---|---|---|---|---|---|---|---|
| **WBES Regular** | Informal Firm Practice | Accss to Finance | Access to Land | Inadequately educated workforce | Electricity | Crime | Difficulty obtaining business permit | Not available |
| | | Foreign exchange | | | Telecommunication | Political instability | Courts | |
| | | Cost of finance | | | Transport | Corruption | Customs | |
| | | | | | Water | Economic Instability | Labor regulation | |
| | | | | | | | The Indiginization & | |
| | | | | | | Uncertain Policy | economic empowerment | |
| | | | | | | | act (Specific to Zimbabwe) | |
| | | | | | | Economic Policy | Zoning restrictions | |
| | | | | | | | Regulation on hours | |
| | | | | | | | of operation | |
| | | | | | | | Regulation on pricing | |
| | | | | | | | and mark-ups | |
| | | | | | | | Tax Rate and Tax Admin | |
| | | | | | | | Customs imports | |
| | | | | | | | Customs exports | |
| | | | | | | | Licensing import | |
| | | | | | | | Tax authority | |
| | | | | | | | Tax compliance | |
| **WBES Micro** | Informal Firm Practice | Accss to Finance | Access to Land | Inadequately educated workforce | Electricity | Crime | Difficulty obtaining business permit | Not available |
| | | | | | Telecommunication | Political instability | Courts | |
| | | | | | Transport | Corruption | Customs | |
| | | | | | | Economic Instability | Labor regulation | |
| | | | | | | | The Indiginization & | |
| | | | | | | | economic empowerment | |
| | | | | | | | act (Specific to Zimbabwe) | |
| | | | | | | | Tax rates and Tax Admin | |
| **WBES Informal** | Informal Firm Practice | Access to finance | Access to Land | Inadequately educated workforce | Electricity | Crime | Crime | Limited Access to Technology |
| | | Cost of finance | | | Telecommunication | Political instability | Courts | |
| | | | | | Transport | Corruption | Customs | |
| | | | | | Water | Economic Instability | Labor regulation | |
| | | | | | | | Tax Rate and Tax Admin | |
| **HHS** | Low Demand for Goods | Access to finance | Unclear Ownership of Land (Only asked in Uganda) | Difficulty recruiting qualified personnel | Electricity | Crime | Legal regulations | Lack of technical management resources |
| | Competition | Access to credit | | Labor cost | Lack of Space or premise for work | Corruption | Tax rates | Lack of technical manufacturing input |
| | Lack of Market Information | Lack of Raw Material | | Limited time available for business | Lack of secure power | Economic Policy | Regulation | |
| | | Lack of Equipment | | | Infrastructure | Insecurity | | |
| | | Low input | | | Poor quality electricity and phone | | | |
| | | | | | Road quality | | | |
| | | | | | Internet | | | |

Table A4: Implied Resource Preference by Survey Type: Alternative Construction

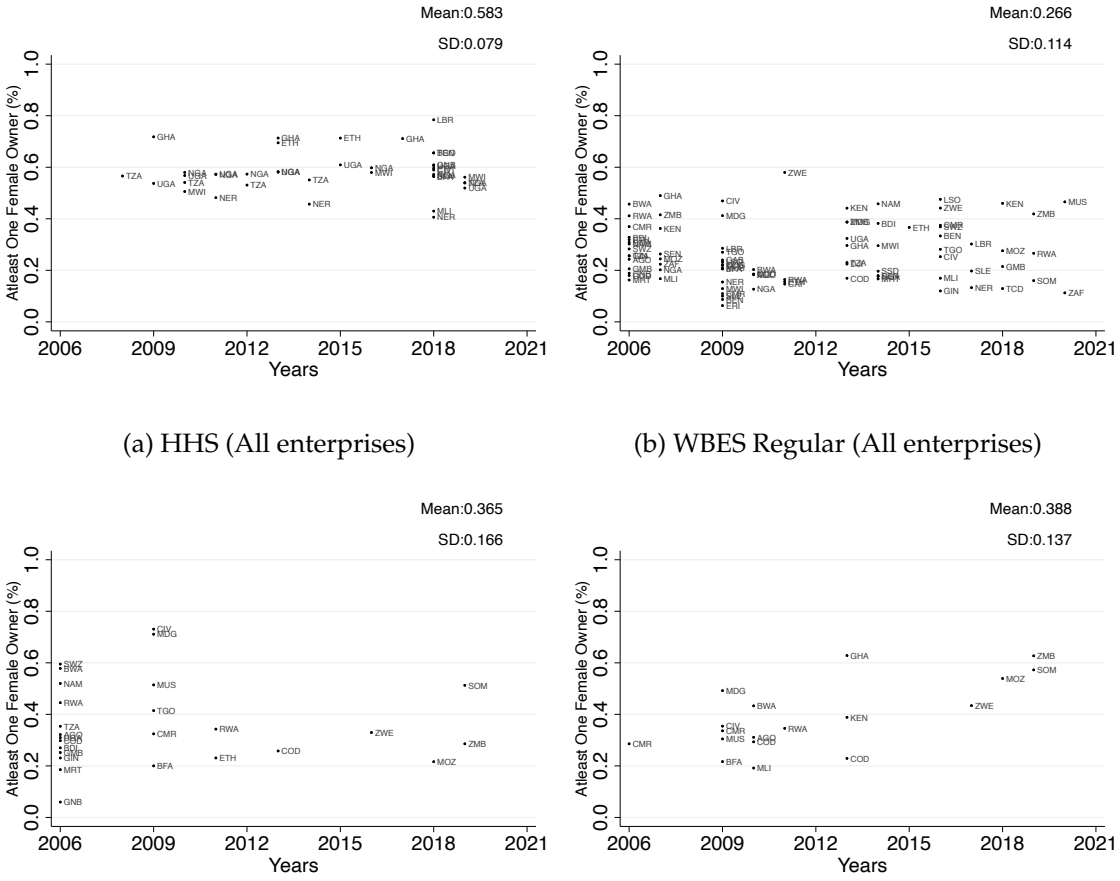
| | (1) Market | (2) Infra. | (3) Capital | (4) Gov. | (5) Safety | (6) Tech. | (7) Labor | (8) Land | (9) Misc. |
|---|---|---|---|---|---|---|---|---|---|
| **Panel A: Original Constraint Responses** | | | | | | | | | |
| WBES Regular | -0.306*** | 0.000 | 0.044 | 0.225*** | 0.057** | -0.014* | 0.007 | 0.054*** | -0.030** |
| | (0.048) | (0.067) | (0.059) | (0.035) | (0.028) | (0.007) | (0.011) | (0.015) | (0.015) |
| WBES Micro | -0.305*** | -0.020 | 0.111* | 0.210*** | 0.055* | -0.013* | -0.008 | 0.061*** | -0.030* |
| | (0.050) | (0.070) | (0.061) | (0.041) | (0.031) | (0.007) | (0.012) | (0.016) | (0.015) |
| WBES Informal | -0.381*** | -0.021 | 0.247*** | 0.047 | 0.061* | 0.000 | 0.004 | 0.115*** | -0.029* |
| | (0.055) | (0.076) | (0.072) | (0.043) | (0.036) | (0.017) | (0.019) | (0.020) | (0.016) |
| HHS Mean | 0.382 | 0.312 | 0.171 | 0.044 | 0.029 | 0.016 | 0.013 | 0.000 | 0.033 |
| R-sqr | 0.820 | 0.744 | 0.693 | 0.756 | 0.727 | 0.382 | 0.517 | 0.717 | 0.529 |
| **Panel B: Randomly Generated Constraint Responses** | | | | | | | | | |
| WBES Regular | -0.075*** | -0.189*** | -0.053*** | 0.361*** | 0.038** | -0.077*** | 0.031* | 0.051*** | -0.063*** |
| | (0.015) | (0.030) | (0.018) | (0.036) | (0.019) | (0.010) | (0.016) | (0.007) | (0.014) |
| WBES Micro | -0.075*** | -0.181*** | -0.046*** | 0.350*** | 0.039* | -0.076*** | 0.045** | 0.053*** | -0.063*** |
| | (0.014) | (0.033) | (0.017) | (0.041) | (0.022) | (0.011) | (0.018) | (0.009) | (0.014) |
| WBES Informal | -0.102*** | -0.111*** | 0.000 | 0.036 | 0.107*** | -0.050*** | 0.095*** | 0.114*** | -0.062*** |
| | (0.019) | (0.039) | (0.021) | (0.046) | (0.025) | (0.017) | (0.031) | (0.012) | (0.015) |
| HHS Mean | 0.157 | 0.320 | 0.142 | 0.101 | 0.082 | 0.087 | 0.046 | 0.007 | 0.056 |
| R-sqr | 0.758 | 0.755 | 0.710 | 0.911 | 0.622 | 0.785 | 0.534 | 0.795 | 0.782 |
| Observations (Panel A and B) | 146 | 146 | 146 | 146 | 146 | 146 | 146 | 146 | 146 |
| **Panel C: P-Value of the Difference in Point Estimates between Panel A and Panel B** | | | | | | | | | |
| WBES Regular | 0.0000 | 0.0001 | 0.0233 | 0.0002 | 0.3670 | 0.0000 | 0.0062 | 0.8318 | 0.0033 |
| WBES Micro | 0.0000 | 0.0012 | 0.0006 | 0.0008 | 0.5163 | 0.0000 | 0.0000 | 0.5716 | 0.0028 |
| WBES Informal | 0.0000 | 0.1031 | 0.0000 | 0.7922 | 0.0837 | 0.0000 | 0.0000 | 0.9613 | 0.0031 |

Note: This table shows the regression of implied resource preferences for different enterprise constraints on survey type. Panel A shows the implied intensity of constraints constructed from original responses whereas Panel B shows the findings from randomly generated constraint responses. Panel C reports the P-value of the difference in point estimates between Panel A and B. Appendix C.7 explains the details of the variable construction. All observations are weighted using their respective survey source sampling weights. HHS is the reference group and each coefficient reflects the difference in means of implied resource preference for the specific constraint between HHS and the respective survey source type. All regressions include fixed effects for the estimates source country and year, respectively to account for time-invariant country characteristics and yearly trends. Robust standard errors are in parentheses.

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

# B Unweighted Appendix Figures

Figure B1: Female Ownership Representation by Data Source, Country and Survey Year




(a) HHS (All enterprises)

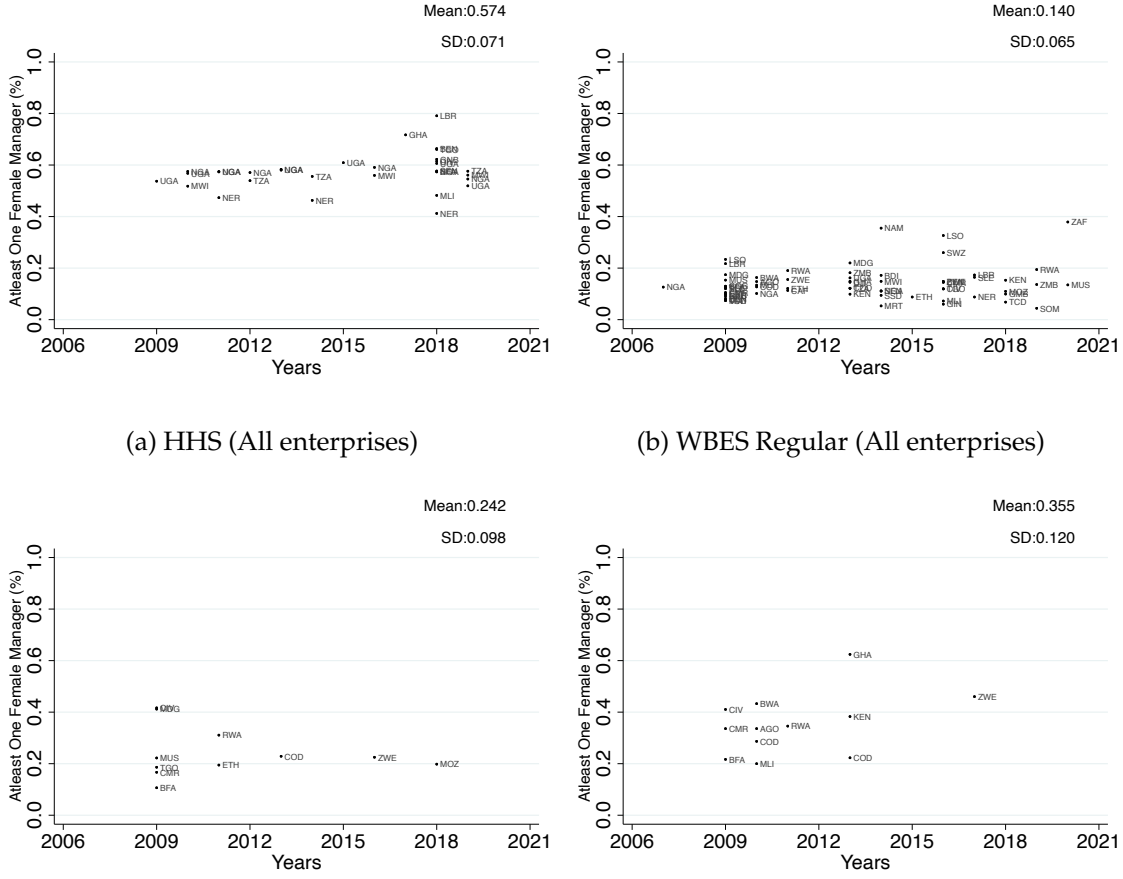
(b) WBES Regular (All enterprises)

(c) WBES Micro Survey

(d) WBES Informal Survey

Note: This figure shows the share of enterprises reporting at least one female owner by country and year of the survey. Figure (a) shows data from 39 Multi-topic Household Surveys (HHS) covering 15 countries. Figure (b) shows data from 85 World Bank Enterprise Surveys (WBES) covering 43 countries. Figure (c) shows data from 26 Micro Enterprise Surveys (WBES Micro) covering 24 countries. Figure (d) shows data from 18 Informal Sector Enterprise Surveys (WBES Informal) covering 15 countries. Observations are not weighted using the survey source sampling weight.

Figure B2: Female Manager/Decision Maker Representation by Data Source, Country and Survey Year




(a) HHS (All enterprises)

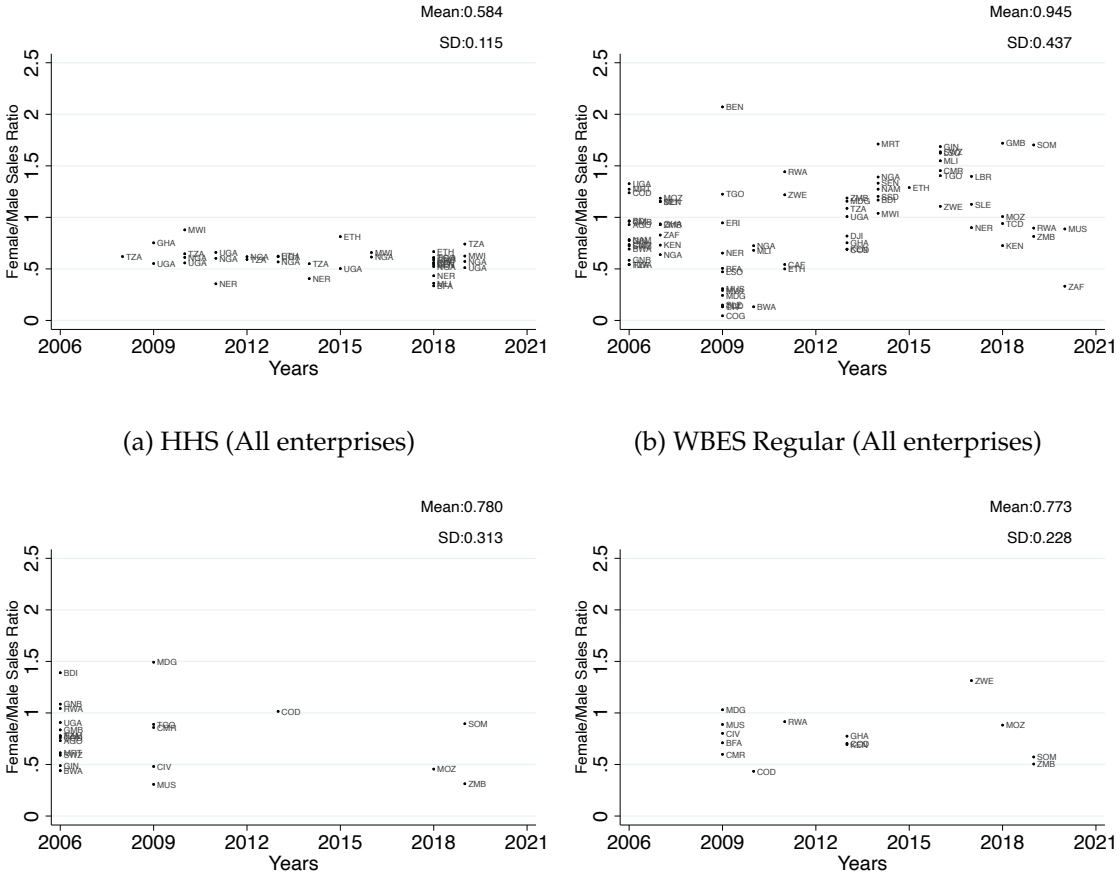
(b) WBES Regular (All enterprises)

(c) WBES Micro Survey

(d) WBES Informal Survey

Note: This figure shows the share of enterprises reporting at least one female manager or decision maker by country and year of the survey. Figure (a) shows data from 38 Multi-topic Household Surveys (HHS) covering 15 countries. Figure (b) shows data from 85 World Bank Enterprise Surveys (WBES) covering 43 countries. Ethiopia is excluded from this analysis (Figure a) because of no information on the female manager. Figure c and d show data from WBES Micro and Informal, respectively. Note that the WBES Micro and WBES Informal have a lot of missing information on the gender of the top manager on decision maker(51.10% and 49.44%, respectively). Observations are not weighted using the survey source sampling weight.

Figure B3: Female to Male Sales Ratio by Data Source, Country and Survey Year




(a) HHS (All enterprises)

(b) WBES Regular (All enterprises)

(c) WBES Micro Survey

(d) WBES Informal Survey

Note: This figure shows the ratio of female- to male-owned enterprise sales by country and year of the survey. Figure (a) shows data for all enterprises from the HHS survey. Figure (b) shows data for all enterprises from the WBES Regular. Figure (c) shows data for all enterprises from the WBES Micro survey. Figure (d) shows data for all enterprises from the WBES Informal survey. In this figure, sales are winsorized at the top 5% level. 18 data sources are excluded from this analysis due to unusually high sales value and uncertainty about the required exchange rate adjustments. Appendix C.3 discusses this in detail. Observations are not weighted using the survey source sampling weight.

# C Data Appendix

## C.1 WBES Informal Sampling Protocol

WBES Informal has employed two different sampling methodologies over time.

The Surveys carried out between 2006 and 2015 involved the creation of different "sampling zones", which are all delineated according to the concentration and geographical dispersion of informal business activity within the same urban centers where the WBES Regular took place that same year. After mapping the sampling zones, enumerators were asked to walk down the two main streets inside each sampling zone from designated starting points and interview one informal manufacturing business and one informal service business along each street. The number of enterprises to be interviewed was fixed and pre-determined, depending on the number of sampling zones mapped.

From 2016 onward, stratified Adaptive Cluster Sampling (ACS) has been used for these surveys. Under this method, first, they divide the urban centers into a grid and select certain squares to be sampled. Then, the enumerators interview every informal business in this area with a short 2-3 minute interview, followed up by a 20-minute interview done on a randomly selected subset of businesses.

## C.2 Construction of Sales Ratio

We estimate female to male-owned enterprise sales ratio to explore how business performance by owner gender varies across survey sources. First, for each data set, enterprise-level sales values are converted into US dollars and adjusted according to the respective PPP exchange rate 2020 recorded for that country and year in the IMF's World Economic Outlook. We winsorized the sales value at the top 95% to deal with possible outliers. Next, we calculate the average sales by owners' gender for each data set. Thus, by dividing the average sales of women-owned enterprises by the average sales of male-owned enterprises, we acquire the female-to-male sales ratio by country, year, and survey.

## C.3 Addendum on Sales Information

A few of the surveys analyzed displayed unusually high sales figures when compared to similar surveys, even after adjusting with their respective exchange rates reported by the IMF's World

Economic Outlook. 12 WBES Regular surveys exhibited mean yearly sales of over USD $10 million, 6 WBES Micro Surveys exhibited median yearly sales of over USD $100,000, 3 WBES Informal countries exhibited median yearly sales of over USD $15,000, and 3 HHS surveys exhibited mean yearly sales of over USD $15,000. After researching the currency history of the countries involved, 6 datasets (5 regular and 1 micro) had their currency re-adjusted to historical currency transitions that were not reflected on the exchange rates. The remaining 18 datasets did not have identifiable solutions to their high sales figures, so they were excluded from all sales-related analyses.

**List of surveys which had their exchange rate re-adjusted:**

- Ghana WBES Regular 2007: 10,000 Cedi → 1 Ghana Cedi
- Mauritania WBES Regular 2006: 10 Ouguiya → 1 Ouguiya
- Mauritania WBES Micro 2006: 10 Ouguiya → 1 Ouguiya
- Mauritania WBES Regular 2014: 10 Ouguiya → 1 Ouguiya
- Zambia WBES Regular 2007: 1,000 Kwacha → 1 Kwacha
- Zambia WBES Regular 2013: 1,000 Kwacha → 1 Kwacha

**List of surveys excluded from sales-related analyses:**

- WBES Regular
  - Angola 2010
  - Benin 2016
  - Cameroon 2009
  - Congo Democratic Republic 2010
  - Cote d'Ivoire 2016
  - Gabon 2009
  - Liberia 2009
- WBES Micro
  - Burkina Faso 2009

  - Ethiopia 2011
  - Kenya 2013
  - Rwanda 2011
  - Zimbabwe 2016
- WBES Informal
  - Angola 2010
  - Botswana 2010
  - Mali 2010
- HHS
  - Ghana 2013
  - Ghana 2017
  - Liberia 2018

### C.4 Alternative Definition of Female Ownership

For each survey source, we construct an indicator representing whether an enterprise has a female manager or not. For the WBES Regular and WBES Micro, we construct the indicator from the question of whether the top manager or main decision maker is female or not. For WBES Informal, the indicator is constructed from the question of whether the main decision-maker is female or not.In the HHS, for each enterprise listed in the non-farm enterprise roster, respondents list the business managers (maximum 2). We use the manager list to determine if there is at least one female manager.

### C.5 Definition of Enterprise Size in Different Surveys

The enterprise size for different surveys is defined as follows:

Appendix Figure A3a: For the WBES Regular, enterprise size is defined by the number of full-term permanent employees. For the LMMIS, enterprise size is defined by the average total number of employees per month.

Appendix Figure A3b: For the WBES Regular and IBES, enterprise size is defined by the total number of permanent employees. For WBES Informal, the total number of paid employees is used to define enterprise size. For the HHS, we use the total number of hired workers (non-household members).

### C.6 Construction of Implied Resource Preference for Dismantling Enterprise Barriers

In this section, we describe how we construct the "Implied Enterprise Resource Preference" variables for each data source from all the different constraint-related questions found in the WBES and HHS surveys. The objective of this variable is to understand, from a policy-maker's point of view, how many resources should be allocated towards solving each of the problems different enterprises are facing. Therefore, we are interpreting the constraint questions as possible signals for a policymaker on which problems are a "priority" by country, year, and survey type.

**Types of Questions Asked:** Overall, there are 5 general types of questions asked across different surveys pertaining to constraints/obstacles faced by the enterprise:

- Which of the following elements of the business environment, if any, currently represents the biggest obstacle faced by this establishment?
  - This question is not open-ended, respondents receive a list of obstacles and they have to choose one among them. The list of available obstacles changes across countries and years
  - Found in WBES Regular, WBES Micro, and WBES Informal for years after 2007, as well as Nigeria HHS and Ethiopia HHS
- Of the problems mentioned above, what are the three most important obstacles for you?
  - This question is not open-ended, respondents receive a list of obstacles and they have to choose three among them in a specific order (from 3rd most important to top importance).

The list of available obstacles changes across countries and years

  - Found in WBES Regular, WBES Micro, and WBES Informal in 2006 and 2007

- Do you think that [blank] presents any obstacle to the current operations of your establishment?
  - Options: 0 “No obstacle”, 1 “Minor Obstacle”, 2 “Moderate obstacle”, 3 “Major obstacle”, 4 “Severe obstacle”
  - This question is usually asked for each of the constraints available as options in the "largest obstacle" question
  - Found in WBES Regular and WBES Micro
- In the last 12 months, has the company encountered the following problems in carrying out its activity?
  - Options: Put 1 for Yes, 2 for No, and 3 for Not concerned/Not applicable
  - Only found on HHS in 2018 with no other constraint question type
- Is [blank] a severe obstacle to the current operation of this business or activity?
  - Options: Put 1 for Yes, 2 for No
  - Only found in WBES Informal

**Interpreting Constraint Responses:** Even if the types of questions above are relatively consistent across surveys, the specific constraints presented to enterprises vary from survey to survey. For example, most WBES Regular surveys done after 2007 ask the question "Which of the following elements of the business environment, if any, currently represents the biggest obstacle faced by this establishment?", and offer the following 15 options to pick their answer from:
Access to finance; Access to land; Business licensing and permits; Corruption; Courts; Crime, theft and disordered regulations; Customs and trade; Electricity; Inadequately educated workforce; Labor regulations; Political instability; Practices of competitors in the informal sector; Tax administration; Tax rates; Transport.

However, if we consider a data set like the WBES Informal Rwanda 2011 data set, the options for this same question are the following: Access to Finance; Access to Land; Corruption; Crime, Theft, and disordered regulations; Electricity; Water.

Differences in possible responses like the ones above are present across many surveys, and many different variations exist for these response options. Given that we want to analyze all this information across all surveys, the different possible constraint options were grouped into 8 different "Broad Categories" that were used in the regression analysis. All the different possible obstacles covered by the different constraint questions, as well as the broad categories they have been allocated to are outlined in Appendix Table A3.

**Standardization and Construction of the Implied Resource Preference:**

Majority of the WBES Regular and Micro surveys ask the categorical constraint questions where the response to a particular business barrier can range from 0 (no obstacle) to 4 (severe obstacle). Whenever available, we use these categorical constraint variables to construct the implied resource reference index because it allows us to capture multiple barriers a business could face at a time while also indicating a degree of severity. For surveys that do not contain this categorical constraint information, but provide binary answers on whether different obstacles are faced by an enterprise or not, we use those binary variables to capture the implied resource preference. Finally, if a survey does not include categorical or binary variables, we use businesses' responses to which is the largest obstacle instead.

Given that a survey could record the enterprise barriers either through categorical variables, binary variables, or variables where enterprises pick their biggest obstacles out of a given list, we standardized these self-reported barrier variables before constructing the broad categories and implied resource preference variables for each data source as follows:

Surveys that include information on categorical variables are standardized by dividing each of the constraint responses by a 'total score', which is obtained by adding the degrees of severity each business has expressed. As a result, responses now reflect a ratio of how much an enterprise reports to be affected by an obstacle in relation to the other available obstacle questions in each survey. These standardized categorical responses are then added together according to their respective broad categories, resulting in 8 variables. We follow the same standardization process for data sources that only include binary response options for constraint questions.

In the case when no other constraint signal is present in a data set except for the answer to the largest obstacle faced by the business, the average response to this question can be calculated on a data set level. By turning the possible responses when selecting their "largest obstacle" into binaries, we can calculate the average rate of response for each option on the obstacle list. This signals how much enterprises would prioritize a solution to a major constraint in relation to the other possible constraint they could pick from in each data set. These standardized responses are then added together according to their respective broad categories, resulting in 9 variables (there are some options in the answer list for this type of question that are not found in the categorical or binary questions).

After this process is complete, all possible constraint questions are compiled into the "Implied Resource Preference" variable. This meets our goal of signaling to a hypothetical policy-maker which are the most significant constraints enterprises face by data set, and how resources should be distributed to solve these obstacles.

**Construction of randomized constraint responses**

As mentioned before, the original questions asked on constraints and a list of possible options varies across data sources. Part of the findings of Table 4 might be driven by these differences in survey construction. To account for this, we create a randomized set of responses to the constraint-related questions for each data source and see how much of the "Implied Resource Preference" can be attributed to the survey construction itself.

Random categorical responses were created by generating a random number between 0 and 1 whenever a business gave a response to a categorical question. The distribution for the random number generation was uniform for the data source the responses belonged to. Given that responses to categorical constraint questions are in a range from 0 to 4, five bands of the same size were assigned to redistribute the responses. If the random number was larger than 0 and smaller than 0.2, the response was 0, if the random number was larger or equal to 0.2 and smaller than 0.4, the response was 1, and so on. In the case of binary variables, the same method described above was used but only two bands were created, one between 0 and 0.5, and another one between 0.5 and 1.

In the case of randomizing top constraint responses, the process begins by generating only 1 random number per enterprise that has indicated their top constraint in their original response. This number is also between 0 and 1, and uniform for the data source the responses belonged to. Then, the bands for each top constraint response were created based on the number of options a business could pick from in a given survey. If a survey could pick between 15 different options as their top constraint, 15 different bands of the same size were created, such that the random number generated above has a 1/15 probability of belonging to a specific band. This process could be easily replicated because luckily the number of questions was co-linear with the options given in a specific survey, so whenever there were exactly 6/8/16 options for enterprises to pick a top constraint from, those options were always the exact same.

### C.7 Construction of Implied Resource Preference: Alternative Approach

The method described above for constructing the "Implied Resource Preference" variable works under the assumption that the categorical obstacle questions are the most informative because that allows an enterprise to inform about the multiple constraints they are facing. However, as a robustness check, we construct the 'Implied Resource Preference' in an alternate way. Here, we first use the "largest obstacle faced by the firm" variable responses whenever available in a data set, followed by the categorical variables, and lastly by the binary variables if the biggest obstacle question is not asked.

The steps for standardizing these variables are exactly the same as described in subsection C.6, with the only difference being the order of operations. We first standardize all available responses to the "largest obstacle faced by the firm", and then whenever this variable is missing, we considered the categorical or binary responses. The results of this robustness check can be found in Appendix Table A4.