WORLD BANK GROUP

GFDRR
Global Facility for Disaster Reduction and Recovery

Resilience Firm Survey

TECHNICAL NOTE

Dar es Salaam, Tanga and Dodoma, Tanzania

1 Overview

Content of this document: This document provides information about The Resilience Firm Survey (RFS). It describes the RFS series: the survey and sampling design and the questionnaire used, and discusses some data considerations, including outlier treatment and anonymization process.

Objective of the survey: The RFS is designed to collect information from private businesses focusing on i) dependence on and reliability of critical and non-critical infrastructure, ii) use of suppliers and impacts of supply chain disruptions, iii) impacts of recent disasters – coping capacity and long-term effects, iv) firm level preparedness and management of shocks and interruptions caused by natural hazards. The survey is customized depending on the context, industry sector and research questions. The data can be used to explore policy-relevant research topics related to climate change adaptation, infrastructure resilience, private sector recovery, and more.

Content of the data: RFS data contain information on firm characteristics, use of suppliers, infrastructure dependency, firm experience with disasters and risk management and coping capacity. Firm characteristics include sector engagement, number of employees, number of clients, costs and sales information. Use of supplier information include location of suppliers, use of inputs, frequency of restocking, storage capacity, etc. Infrastructure dependency focuses on use of infrastructure, such as water, electricity and transport, frequency of disruptions, impacts of disruptions on demand and sales, use of backup infrastructure, etc. Firm experience with disasters captures both direct (damages to property) and indirect (infrastructure and supply chain disruptions) impacts of recent disasters, as well as coping behaviors and recovery of sales after a shock. Risk management and coping capacity capture information on use of hazard risk information, access to insurance and preparedness measures. The RFS can be customized to collect information on different sectors and type of disasters. So far, it has focused on the impacts of urban flooding (TZ) and hurricanes/storms (CAR).

Tanzanian application: The RFS In Tanzania was conducted in three major cities, Dar es Salaam, Tanga and Dodoma, between September and November 2018. The survey focused on firms in manufacturing and service industries and the data was collected via in-person interviews. The survey focused on how firms are affected by floods and which strategies they use to cope and adapt, Impact of flooding considers both direct damages and indirect effects through infrastructure systems, supply chains and workers.

This project was a collaborative effort between Global Facility for Disaster Reduction and Recovery (GFDRR) and Urban, Disaster Risk Management, Resilience and Land Global Practice (GPURL).

Data files and other resources

- RFS_Tanzania_2018: Data in STATA format (.dta)
- RFS_Tanzania_2018_Questionnaire: Questionnaire in Common Application Programming Interface (CAPI) (.pdf)

Citation requirements:

The World Bank. Resilience Firms Survey (RFS), Dar es Salaam, Tanga and Dodoma, Tanzania, 2018. Dataset downloaded from microdata.worldbank.org on [date].

2 The survey

2.1 Description

Name of the study: Resilience Firm Survey (RFS) Tanzania

Geographical coverage: Dar es Salaam, Tanga and Dodoma

Number of observations: 837 firms

Date of the survey: September to November 2018

Primary Investigators: Jun Rentschler (World Bank), Julie Rozenberg (World Bank), Alvina Erman (World Bank), Stephane Hallegatte (World Bank)

Collaborators: GFDRR, GPURL and UDA consulting

Funding: GFDRR

Related reports: Rentschler, et al. (2021)

2.2 Sampling design

The sampling frame consisted of the official business registry obtained from the National Bureau of Statistics' (NBS) in Tanzania. It contained a list of 58,959 firms in Dar es Salaam, Tanga, and Dodoma. Firms listed without a functioning telephone number were excluded from the sampling frame.

All registered firms were divided into five distinct strata, depending on their reliance on transport systems. Ordered from low to high transport reliance, these strata contain firms from the following sectors:

1 Accommodation and food service activities, Construction

2 Communication and other services

3 Manufacturing, Mining, Water, Energy, Agriculture

4 Transportation and storage

5 Wholesale and retail trade, Repair of motor vehicles and motorcycles

The sample selection was completed in one stage, with firms selected through a systematic random sampling method from each stratum. In terms of regional distribution, the sample contains 623 firms from Dar es Salaam, 101 from Dodoma, and 113 from Tanga. About 10 percent of listed firms were interviewed, with the main reasons for the low response rate being incorrect addresses, expectation of payment for survey participation, and concerns about disclosure of tax-relevant information.

All survey results, using the survey weights computed, should be interpreted as representative of private businesses located in the 3 cities Dar es Salaam, Dodoma, and Tanga and registered in the NBS business registry with phone access.

3 Questionnaire modules

The survey modules are the following:

- Firm characteristics
    - Sector
    - Number of workers
- Infrastructure dependence and disruptions
    - Water
    - Electricity
    - Communication (phone and internet)
- Suppliers
- Workers and transport disruptions

- Clients
- Most serious natural shock that affected the firm
- Operational costs
- Investments

4 Data considerations

4.1 Anonymization of the dataset

Protecting the privacy of survey respondents is of the outmost importance to the World Bank. To make sure the data cannot be used to identify individual households in the dataset, a technique of statistical disclosure control (SDC), as described in Benschop et al. (2021), was applied. It helped identify variables that included unique information about households. After identifying the high-risk variables, necessary adjustments were made to make sure the SDC analysis provided satisfactory results, i.e., low risk of re-identification. Results can be shared upon request. The following data editing was done for anonymization purposes:

- Precise location data, such as GPS coordinates
- Identifying and contact information, such as firm name, respondent's name, phone numbers and email contacts, were dropped

4.2 Outlier treatment

Continuous variables may present some measurement errors. A technique of outlier treatment is recommended. Some of these variables are:

- *q3_4_2*: How many KWH do you use to maintain this firm? (in *KWH*)
- *q9_1*: What is your total operational monthly cost? (in *TSh*)

An established method to identify outliers is to tag the observations that deviate from the mean by a set number of standard deviations. Three standard deviations are commonly used. Figure 1 includes STATA code that can be used to tag outliers[1].

Figure 1: Codes for the identification of outliers for the variable *q9_1*

[1] Additional checks may be conducted to analyze the presence of outliers. The technique in Figure 1 assumes that the distribution of the variable is normal. This may not be the case, even after using a logarithmic transformation. Other transformations and for which kinds of variables to use them are explained in Ravallion (2017). Outliers may influence the mean and the median of the distribution. More robust methods of outlier treatment may be necessary, for instance, the median absolute deviations (MAD) method (Belotti et al., 2021, Rousseeuwand Croux, 1993).

```
* Create dummy for outliers
foreach var of varlist q9_2 {
    quietly summarize `var'
    g Z_`var'=(`var' > 3*r(sd)) if `var'<.
    list `var' Z_`var' if Z_`var'==1

}
/*
        +---------------------+
        |       q9_2   Z_q9_2 |
        |---------------------|
     49. |  800000000        1 |
    565. |  397000000        1 |
    633. |  296334635        1 |
    699. | 1700000000        1 |
    752. | 2000000000        1 |
         +---------------------+

*/
```

These outliers can then be either dropped or replaced. For instance, for the variable *q9_1*, outliers could be replaced by a predicted value based on characteristics associated with operational costs in the sample (excl, outliers), such as number of clients, sector, number of employees, etc. This technique is linked to high uncertainty and careful consideration should be taken before using predicted data.

References

BELOTTI, F., G., VECCHI, G., AND MANCINI (2021): “OUTDETECT: Stata module to perform outlier detection and diagnostics for welfare analysts,” *Statistical Software Components* S458932, Boston College Department of Economics.

BENSCHOP, T. AND WELCH, M. (n.d.): “Statistical Disclosure Control for Microdata: A Practice Guide”, Retrieved 30 March, 2022, from https://sdcpractice.readthedocs.io/en/latest/

RENTSCHLER, J. E., KIM, E. J., THIES, S. F., DE VRIES ROBBE, S. A., & ERMAN, A. E. (2021). Floods and Their Impacts on Firms: Evidence from Tanzania. Policy Research Working Paper. The World Bank. Washington DC.

RAVALLION, M. (2017): “A concave log-like transformation allowing non-positive values,” *Economics Letters*, 161, 130-132.

ROUSSEEUW, P. J., CROUX, C. (1993): “Alternatives to the median absolute deviation,” *Journal of the American Statistical association*, 88(424), 1273-1283.