VAT Cashback Programs in Practice - the Case of *Devolve-ICMS* program in Rio Grande do Sul, Brazil

Code Repository Documentation

Pierre Bachas

Tatiana Flores

Gabriel Lara Ibarra

Anderson Mantovani

Evandro Costa Souza de Oliveira

Giovanni Padilha da Silva

Thiago Scot*

*Corresponding author: Thiago Scot. Email: tscot@worldbank.org

Contents

1   Overview

The codes in this repository replicate the tables and figures from "VAT Cashback Programs in Practice - the Case of *Devolve-ICMS* program in Rio Grande do Sul, Brazil", by Bachas, Flores, Ibarra, Mantovani, Oliveira, Padilha and Scot. The replication folder contains the codes to go from the raw administrative data to the results in the paper.

This documentation is structured as follows. Section 2 describes the data sources and their availability. Section 3 describes the datasets used in the analysis. Section 4 details the computational requirements. Section 5 provides instructions to replicators. Section 6 provides a mapping between the codes and the tables and figures of the paper. Finally, section 7 describes the scripts used for the replication.

2   Data Availability and Provenance Statements

☐ This paper does not involve analysis of external data (i.e., no data are used or only data are generated by the authors via simulation in their code).

2.1   Statement about Rights

☑ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

2.2   Summary of Availability

☐ All data **are** publicly available.

☐ Some data **cannot be made** publicly available.

☑ **No data can be made** publicly available.

The administrative data for this project is owned by the Rio Grande do Sul Finance Secretariat (SEFAZ-RS). The data were made available to us exclusively for this research project through collaboration agreements between World Bank authors and the SEFAZ-RS.

Individuals interested in accessing the data for replication purposes can contact Thiago Scot (`tscot@worldbank.org`). We will provide reasonable assistance to requests for clarification and replication, and will support application for data access through an agreement with SEFAZ-RS.

Researchers interested in obtaining the data for their own analyses can directly contact SEFAZ-RS[1].

2.3 Details on each Data Source

The analysis is based on de-identified administrative raw data provided by the SEFAZ-RS to the research team on June 2024. Raw data includes:

- **Household information.** This dataset includes the most updated data on beneficiaries from the *Devolve-ICMS* program. Part of the data is sourced from the Brazilian national register of social assistance (*Cadastro Único*) and includes information such as self-reported income, family size, municipality of residence and gender of household head. The dataset also includes other information on households' program participation, such as date of entry into the program and whether the family has picked up the debit card provided.

- **Monthly program information.** This is the mian dataset used to define households' benefits. The main source of information for transfers calculation is the monthly consumption linked to household head. All other variables in this dataset are formulaic, and are calculated based on consumption and other household characteristics, to derive program benefits.

---

1. Giovanni Padilha (GiovanniS@sefaz.rs.gov.br).

3 Datasets

This section describes every dataset used to replicate the results. Table 1 displays the set of raw datasets (found in `Raw` data folder of the replication package) needed for full replication from scratch. Table 2 list the set of processed datasets (generated by code and stored in the `Intermediary` folder of the replication package).

Table 1: Raw datasets

Dataset	Description	Notes	Provided
Devolve_WB_Moradores_20240626_Hash.csv	Household level dataset	Confidential	No
Devolve_WB_20230423_Hash.csv	Household-month level dataset, for program implementation	Confidential	No

Table 2: Processed datasets

Dataset	Description	Notes	Provided
data_fixed.RDS	Clean version of household level data	Confidential	No
data_panel.RDS	Cleaned version of monthly panel data	Confidential	No

4 Computational Requirements

4.1 Software requirements

Required software is `R`. The code was run in R version 4.4.2 (2024-10-31).

4.2 Memory and runtime requirements

The code was run for the last time in a Macbook Air M2, 2022, 24 GB RAM. A replicator could expect the whole code to run in ≈ 1 minute with those settings.

5 Instructions to replicators

5.1 Folder structure

For full replication of the project, we suggest the following folder structure.

```
Data
        – Raw
        – Intermediate
Code
Outputs
        – Graphs
        – Tables
```

In the master do-file, we set the `paths` so the above folder tree can run easily. The codes described in section 7 rely on this folder structure.

6 List of tables, and figures

The provided code reproduces:

☐ All numbers provided in text in the paper

☑ All tables and figures in the paper

☐ Selected tables and figures in the paper, as explained and justified below.

6.1 Mapping of tables and analysis code

Table 3 provides a mapping between all tables of the paper and the codes producing these results. The codes generate .tex files containing the tables.

Table 3: Mapping of tables and analysis code

Exhibit name	Output filename	Code
Table 1	table_reg_1	2_Devolve_ReportResults.qmd
Table 2	table_reg_2	2_Devolve_ReportResults.qmd

6.2 Mapping of figures and analysis code

Table 4 provides a mapping between all figures of the paper (including the online appendix) and the codes producing these results. The codes generate files in .pdf format containing the graphs.

Table 4: Mapping of figures and analysis code

Exhibit Name	Panel	Output filename	Code
Figure 1		consumption_month.png	2_Devolve_ReportResults.qmd
Figure 2a		consumption_mean_distribution.pdf	2_Devolve_ReportResults.qmd
Figure 2b		consumption_quantiles_distribution.png	2_Devolve_ReportResults.qmd
Figure 3		consumption_incomegroups.png	2_Devolve_ReportResults.qmd
Figure 4		nfg_formal_consumption.png	2_Devolve_ReportResults.qmd

7 Description of programs and code

The project only includes two scripts, other than the master script. The first script cleans the raw data, while the second script is a Quarto file that generates a report with all figures and tables.

Global paths to the different folders of the project (input files, output tables, figures, etc.) are established in the master do-file. The replicator can centrally adjust the links in these files without adjusting the (relative) paths in the specific cleaning or analysis parts. Moreover, the master do-file installs all packages needed in the project.

7.1 Programs

Our code utilizes the following commands/packages installed through the master do-file:

- `pacman`
- `stringr`
- `stringi`
- `dplyr`
- `data.table`
- `ggplot2`
- `lubridate`
- `tictoc`