

Who benefits from food tax exemptions in lower-income settings? Evidence on pass-through and incidence (Papua New Guinea)

Replication package (code and documentation). This repository holds the code that produces every exhibit in the paper and appendix. Data are not included; see [Data availability](#).

Contents

1. [Overview](#)
2. [Data availability](#)
3. [Instructions for replicators](#)
4. [List of exhibits](#)
5. [Requirements](#)
6. [Code description](#)
7. [Folder structure](#)

Overview

The analysis runs in three stages: Stata prepares the data, R builds two maps, then Stata runs the analysis, tables, and figures. Two master scripts drive it:

- `main.do` (Stata): sets one path, installs packages, builds the `Final/` datasets, pauses for the R stage, then runs all analysis.
- `main.R` (R): sets one path, builds Figures 1 and 2, and writes two intermediate datasets the Stata analysis reads.

A reviewer runs the whole package after changing one line in each master script (the top-level directory). Nothing else needs editing.

Data availability

Summary classification: some data cannot be made publicly available. This repository ships code and documentation only. No raw or intermediate data is included. Before running, place the data under a `Data/` folder matching the paths in `main.do` Section 1.

Data collected by the authors (Sources 2, 4, and 5) are archived in the World Bank Microdata Library and are publicly accessible at the links below, in line with World Bank policy on data created for Bank projects.

#	Source	Provider	Access	Link or contact	Accessed
1	Phone	World Bank	Restricted,	https://microdata.pacific	2026-04

#	Source	Provider	Access	Link or contact	Accessed
	survey, household food module (PNG_HFPS_Household_weighted_wFood.data)	High-Frequency Phone Survey, Papua New Guinea	licensed	data.org/index.php/catalog/877	
2	Expert prediction survey (experts_survey.csv)	Authors, fielded via Qualtrics	Public	Forthcoming MICRODATA LIBRARY	2025-07 to 2025-09
3	2011 census population by district (png_admpop_adm2_2011_v2.csv)	PNG National Statistical Office, distributed via HDX	Public	https://data.humdata.org/dataset/cod-ps-png	
4	Store price collection, Port Moresby (POM PRICE COLLECTION.xlsx)	Authors, field collection	Public	Forthcoming MICRODATA LIBRARY	2025-05 to 2025-10
5	Online store prices (foodpro/*.csv, rh/*.csv)	Authors, web scraping of rhtradingpng.com and fpr.com.pg	Public	Forthcoming MICRODATA LIBRARY	2025-05 to 2025-08
6	Retail Price Index (World_Bank__RPI_Data_Reque	Bank of Papua New Guinea	Restricted, not redistributable	Data request to ilibitino@bankpng.gov.pg	2026-05

#	Source	Provider	Access	Link or contact	Accessed
7	Administrative prices and quarterly CPI (Comparison_Group.xlsx, GST exempt goods WB tracking - Edited.xlsx, Table_13_December_Qtr_2025.xlsx)	PNG National Statistical Office	Restricted, except the published Table 13 CPI release	Data request to barry.lei@nso.gov.pg ; CPI release at https://www.nso.gov.pg/statistics/economy/consumer-price-index	2026-04
8	Business Sentiment Survey (Copy_of_Wholesale_retail_competition.xlsx, World_Bank_request.xlsx)	Bank of Papua New Guinea Business Sentiment Survey	Restricted, not redistributable	Data request to ilibitino@bankpng.gov.pg	2026-07

Notes on individual sources

Source 1. The extract used here sits in a dated subfolder (20260429) and is derived from the Pacific Data Hub catalog entry above, reference SPC_PNG_2023_HFPS-Q2.

Sources 2 and 4. Direct identifiers were removed before archiving: respondent email address, IP address, and geolocation in the expert survey; enumerator names in the store price workbook.

Source 5. The scrapers (4_online_appendix/web_scraping/) require a manual browser verification step mid-run, so they are a one-off collection tool, not an automated pipeline step. The exact snapshots cannot be re-collected: the live sites change over time and prior pages are not archived. The archived files contain prices, standardized item codes, and

collection dates; verbatim product text and URLs are excluded. The dated CSVs feed `build_online_prices.do`.

Restrictions on access, publication, and retention

Sources 1, 6, 7, and 8 are not redistributable under the terms set by the data providers. Only derived, non-identifying results appear in the paper. Sources 2, 4, and 5 are released without restriction through the Microdata Library.

Rights statement

The authors have legitimate access to and permission to use all data used in the manuscript.

Instructions for replicators

1. Obtain the data (see the availability statement) and place the files under `Data/Raw/...` following the folder names in `main.do` Section 1 (census, experts_survey, phone_survey, store_collection, e_store_collection, rpi, nso, sentiment). The phone survey extract sits in a dated subfolder (20260429).
2. Install the software and packages in [Requirements](#).
3. Open `main.do`. Set the top-level directory on the single line marked `CHANGE ONLY THIS ONE LINE`:
 - `main.do`, line 37: `global root "..."`
4. Run `main.do`. It prepares the data and pauses with an on-screen instruction. The pause requires interactive Stata; running `main.do` in batch mode will not stop at the R stage.
5. Open `main.R`. Set the same directory:
 - `main.R`, line 26: `root <- "..."` Run `main.R` fully. It writes Figures 1 and 2 and two intermediate datasets.
6. Return to `main.do` and press any key. It runs the analysis and writes all tables and figures to `Outputs/`.

The run order is fixed by the master scripts. Data-prep files are numbered in run order (01–10). One dependency inside the analysis: `fig4ab_event_study_competition.do` writes `store_competition_groups.dta`, which `fig4c_within_chain.do` reads; `main.do` already runs them in that order.

List of exhibits

Every table and figure, mapped to its output file, script, and export line, is in `docs/exhibit_manifest.md`. Figure filenames do not always match the exhibit number in the paper (for example Figure 4d exports to the `event_study/` subfolder); the manifest gives the exact mapping.

The code reproduces the tables and figures listed in the manifest. Exhibits in the manuscript that do not appear there (for example a descriptive Table 1 or Table A1, and appendix figures A8, A9, A16) are not generated by this package; they are built by hand or excluded by design. See the Notes section of the manifest.

docs/exhibits.tex compiles all figures and tables in paper order and is the only copy of that file.

Requirements

Software

- **Stata 17.** User-written packages (installed automatically by main.do Section 3 via `ssc install`): `estout`, `gtools`, `winsor2`, `reghdfe`, `ftools`, `regsave`, `mmat2tex`, `boottest`, `distinct`.
- **R 4.6.** Packages (installed automatically by main.R if missing): `tidyverse`, `sf`, `ggplot2`, `dplyr`, `readr`, `haven`, `viridis`, `scales`, `RColorBrewer`, `cowplot`, `knitr`, `readxl`, `data.table`, `osmdata`.
- **Python 3.11** (only to re-run the online-price scrapers, optional): `selenium`, `undetected_chromedriver`, `certifi`, plus the local helper `4_online_appendix/web_scraping/utils.py`. The scrapers need a manual Cloudflare step and are not part of the automated run.

Runtime and storage

Runtime depends on the machine; the phone-survey inline rebuilds and the household-cluster bootstraps (500 replications in `fig5_incidence.do` and `fig6_actual_vs_expert.do`) are the slowest steps. Allow storage for the Final/ datasets and Outputs/ beyond the code footprint.

Code description

```
main.do    Stata master: paths, packages, data prep, pause for R, analysis
main.R     R master: builds Figures 1 and 2, writes intermediates for Stata

1_data_prep/    Builds Final/ datasets from Raw/ (run order 01-10)
2_maps/         R maps (Figures 1 and 2) and store proximity tertiles
3_analysis/     Tables and figures (named by exhibit)
4_online_appendix/ Online-price cleaning, Figure 4d, and the scrapers
5_experts/      Expert prediction survey figures and Table A5
docs/          Exhibit manifest and exhibits.tex (compiles all figures/tables
               in paper order)
```

Data-prep files are numbered in run order. Analysis files are named by the exhibit they produce (for example `table2_passthrough.do`, `fig5_incidence.do`, `figA19_event_study_hh.do`), so the manifest maps cleanly to filenames. Each script carries a header block stating its purpose, inputs, outputs, and dependencies.

Folder structure

The full package (assembled on the author's machine, not shipped here) has this layout. The root set in the master scripts points at its top.

```
root/
├── Code/           (this repository)
├── Data/
│   ├── Raw/       (census, experts_survey, phone_survey, store_collection,
│   │               e_store_collection, rpi, nso, sentiment)
│   └── Final/     (built by 1_data_prep and 2_maps)
├── Outputs/
│   ├── Figures/
│   └── Tables/
└── logs/
```

Data/, Outputs/, and logs/ are excluded from version control (see .gitignore).