

README: Reproducibility Package

Expanding pension coverage to the informal sector: the case of Viet Nam

By Trang Luu, Nga Thi Nguyen, and Robert Palacios

Contents

1. [Overview](#)
2. [Data Availability](#)
3. [Instructions for Replicators](#)
4. [List of Exhibits](#)
5. [Requirements](#)

Overview

This README file lists the data sources and codes used for the analysis of the results included in the paper “Expanding pension coverage to the informal sector: the case of Viet Nam”. The Tables and Figures included in the manuscript can be replicated by running the two Stata do-files to i) clean the raw data and ii) run the analysis.

Data Availability

Some data supporting the findings of the study cannot be made publicly available. Public aggregated data at the province-level can be downloaded from the Vietnam General Statistic Office website (<https://www.gso.gov.vn/en/statistic-book/>). The Vietnam Labor Force Survey and Vietnam Social Security Administrative Data are not publicly available. Users who would like to replicate the results should contact GSO and VSS to acquire access to such data.

Data Sources

Some data used for the analysis are publicly available and can be downloaded from institutions’ websites. Other data sources are not publicly available and users are recommended to contact GSO and/or VSS to acquire access to such data.

Data	Source	Access/URL
Vietnam Labor Force Survey	General Statistic Office	Not publicly available. Interested users can contact GSO for access. Variables included: IDHO

		C14C	C32	C60	C68
		MaTinh C61	C14D C69	C33	
		MaHuyen C62	C14E HDKT19	C35B	
		MaXa C62A1	C14F Final_Weight	C37B	
		MaDiaBan C62A2	C15B tuoid	C38	
		HoSo C62A3	C15C TTNT	C42	
		ThangDT C62A5	C16A REGION6	C55	
		IDTV C62A6	C16B Region8	C39	
		NamDT C62A4	C16C SEX	C40	
		C01 C62A7	C17 NHOMTUOI3	C40A	
		C02 C62A8	C18 INDKEY	C41	
		C03A C62B	C19 TSTT	C43	
		C03B C62C1	C20 SOGIOLAMVI~3	C44	
		C04 C62C2	C21 Thieuviecl~1	C45	
		C05 C62C3	C21K nganhkt1	C46	
		C06A C62C4	C22 TONGTHUNHAP	C47	
		C06B C62C5	C23 nganhkt	C47A	
		C07 C62C6	C24 CMKT	C47B	

		C08 C25 C47C C62C7 nhomtuoi_5 C09 C25K C48 C62D1 LLLD C09A C26 C50B C62D2 NHOMTUOI4 C09B C27 C52B C62D3 nhomtuoi_8 C10 C28 C53 C62D4 nghe C11 C29 C54 C63 QUADAOTAO C12 C30 C56 C64 C13 C30A C57 C65 C14A c31 C58 C66 C14B C31K C59 C67
Social insurance administrative data	Vietnam Social Security	Not publicly available. Interested users can contact VSS for access.
PAPI index	UNDP	https://papi.org.vn
Province-level income and poverty index	General Statistic Office – Statistical Yearbook (2021)	https://www.gso.gov.vn/en/data-and-statistics/2022/08/statistical-yearbook-of-2021/
Number of contributors to China's Rural and Urban pension scheme	China's Ministry of Human Resources and Social Security (MOHRSS)	https://www.mohrss.gov.cn/ and — Statistical Bulletins / 年度统计公报 and the China Statistical Yearbook / China Social Insurance Statistical Yearbook (social security chapter), where the row labelled 当年缴费人数 / 缴费人数 appears for the rural / new-rural /urban-rural resident pension for different years.

Instructions for Replicators

New users should follow these steps to run the package successfully: - Users must first have access to all data files if they are not included in the reproducibility package. They should go to the mentioned links, download the listed files, and place them in the data folder.

- 01_data_cleaning.do
- Run this do-file to clean raw data files and prepare working files for the analysis
- 02_analysis.do
- Run this do-file for the whole analysis included in the manuscript

List of Exhibits

Given users' access to all the raw data files, the provided code reproduces:

- All numbers provided in text in the paper
- All tables and figures in the paper

The order of exhibit shown below is as appears in the manuscript:

Exhibit name	Output filename	Script	Note
Figure 1	VSIF_figure1_2_3.xlsx (sheet "Figure 1")	Manually created in Excel	Found in Outputs/main/figures (<i>raw data not publicly available</i>) The number of mandatory and voluntary contributors were provided by the Vietnam Social Security (VSS) based on administrative data. The labor force is also provided by VSS and refers to the population estimated to be aged 15 up to the retirement age, 60 for males and 55 for females using National Statistical Office figures
Figure 2	VSIF_figure1_2_3.xlsx (sheet "Figure 2")	Manually created in Excel	Found in Outputs/main/figures (<i>raw data not publicly available</i>) The figure uses administrative data provided by the VSS. The mandatory and voluntary contributors by age were summed to arrive at the totals shown in the figure. The error in the

Exhibit name	Output filename	Script	Note
			sheet has been corrected.
Figure 3	VSIF_figure1_2_3.xlsx (sheet "Figure 3")	Manually created in Excel	Found in Outputs/main/figures (<i>raw data not publicly available</i>) The figure uses administrative data from VSS on contributions collected (in billions) and the number of voluntary contributors per year in thousands
Figure 5	Figure 5 china_rural_pension_contributors.xlsx (sheet 1)	Manually created in Excel	Found in Outputs/main/figures (<i>raw data not publicly available</i>) The figure uses administrative data for the number of contributors to China's Rural and Urban pension scheme. Data come from official government statistics published by the Ministry of Human Resources and Social Security (MOHRSS). See https://www.mohrss.gov.cn/ and — Statistical Bulletins / 年度统计公报 and the China Statistical Yearbook / China Social Insurance Statistical Yearbook (social security chapter), where the row labelled 当年缴费人数 / 缴费人数 appears for the rural / new-rural /urban-rural resident pension for different years.
Table 4	SI_factors.xlsx	02_analysis.do (lines 30-65)	Found in Outputs/main/tables Table format: label column names, label names of categorical variables
Figure 6	SI_factors_byage.png	02_analysis.do (lines 68-83)	Found in Outputs/main/figures
Figure 7	VSIF_income_corr.xlsx	02_analysis.do (lines 89-104)	Found in Outputs/main/tables Panel A: insert scatter plot, using columns monthly_inc_pc and shr_vsif Panel B: insert scatter plot, using columns poor and shr_vsif Add axis titles and trend lines for both graphs

Exhibit name	Output filename	Script	Note
Figure 8 Panel A	contributor.png	02_analysis.do (lines 107-121)	Found in Outputs/main/figures
Figure 8 Panel B	Shrcontributor.png	02_analysis.do (lines 123-132)	Found in Outputs/main/figures
Table 5	reg_matching.xlsx	02_analysis.do (lines 137-182)	Found in Outputs/main/tables Table format: label column names, label names of standardized variables
Figure 9	VSIF_projection.xlsx	02_analysis.do (lines 200-240)	Found in Outputs/main/tables Insert bar and line chart for both graphs, add labels.
Table 6	reg_matchingpoor.xlsx	02_analysis.do (lines 184-195)	Found in Outputs/annex/tables

Requirements

Software Requirements

- **Stata version 17**
 - outreg2

Memory and Runtime and Storage Requirements

The codes run file in a normal computer memory setup. Runtime takes approximately 3 minutes for each do-file.