

**README for the Reproducibility Package for
The World Bank Policy Research Working Paper
“ Intangible Capital and Labor Income: Firm-Level Evidence from India ”**

Abdul A Erumban*, Gaurav Nayyar**, and Shu Yu**

*Groningen Growth and Development Center, University of Groningen.

**World Bank

August-12, 2026

Table of Content

Reproduction Instructions.....	1
Data Statement.....	2
List of Exhibits.....	4

Reproduction Instructions

To run the scripts, new users need to change the directory path on line 7 of the "Master_Active.do" file. After changing the directory, running the “Master_Active.do” file will produce all the tables and figures as they appear in the manuscript by their corresponding numbers.

The global paths are set for do files, input data, output tables, and output figures, respectively, as follows (see lines 9-12 in the “Master_Active.do”:

- global dofiles \$path\2 - Analysis\1 - Do-Files
- global input_data \$path\1 - Data
- global output_table \$path\2 - Analysis\2 - Outputs\1 - Tables
- global output_figure \$path\2 - Analysis\2 - Outputs\2 – Figures

***Shu Yu (syu2@worldbank.org; corresponding author) is a Senior Economist with the Chief Economist of the Prosperity Vertical of the World Bank. **Gaurav Nayyar (gnayyar@worldbank.org) is a Lead Economist with the DEC of the World Bank. * Abdul A. Erumban (a.a.erumban@rug.nl) is an Assistant Professor at the University of Groningen.

Acknowledgements. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. The authors are grateful to Aart Kraay, Bart van Ark, Carol Corrado, Daniel Lederman, Massimiliano Iommi, Filippo Bontadini, Bishawanath Goldar, Taehyoung Cho, Hak K. Pyo, and Keun Rhee for their comments at various stages of this paper. Dana Vorisek contributed to the project during the proposal stage. We thank Suresh Aggarwal for his help in extracting and preparing the Labor Force Survey data. We also appreciate the insights and suggestions from participants at the World KLEMS Conference in Tokyo, the Louiss- WIPO Global INTAN-Invest Conference in Rome, the seminar at the Bank of Korea, the 2025 Productivity Research Conference in Manchester, and the SEM Conference in Athens. We acknowledge the financial assistance provided by the World Bank to Abdul Erumban for this project.

Data Statement

All publicly available data are provided in ready-to-use files in the reproducibility package.

Dataset	Located at	Data URL and additional information	Date acquired
Sector taxonomies	\$input_data\\taxnomies.dta	Skill intensity (India KLEMS July 2024 version & Labor force surveys); R&D and ICT intensities (Prowess). Processed using « \$dofiles\\SkillShare_Taxonomy » on the following two databases: “\$input_data\\KLEMS_stata.dta “ and “\$input_data\\EmpWGbyEDU.dta”(cannot be shared on public domain).	Jan 2025
Final regression dataset	\$output_data\\IND_intan_regData_v5.dta	Processed using "\$dofiles\\Replication_IND_intan_regData_v5.do", raw data is downloaded from the Prowess database (not publicly available, only available to license holders). Please check the do file and below for details.	April 2023

ADDITIONAL INFORMATION FOR \$INPUT_DATA\\TAXNOMIES.DTA

The skill intensity measures in this dta file was derived from [India KLEMS](#) July 2024 version and [Labor force surveys](#). \$input_data\\taxnomies.dta is generated by running « \$dofiles\\SkillShare_Taxonomy » on the following two databases: \$input_data\\KLEMS_stata.dta and \$input_data\\EmpWGbyEDU.dta (cannot be shared on public domain).

1. **Filename(s):** \$input_data\\KLEMS_stata.dta

(data downloaded from the RBI KLEMS website, see below)

[users can directly access the file from the RBI website, and therefore, it may not be shared on public domain]

- **Source:** Reserve Bank of India (2024)

License terms and access details. The India KLEMS July 2024 database is publicly available from the Reserve Bank of India (<https://rbi.org.in/Scripts/KLEMS.aspx>), subject to the applicable terms of use.

- **Accessed** in: July 2024
- **Citation:** Reserve Bank of India (2024). *India KLEMS Database, July 2024 Release.*, <https://rbidocs.rbi.org.in/rdocs/content/docs/INDIAKLEMS08072024.xlsx>
- Das, D. K., Erumban, A. A., Aggarwal, S., & Das, P. C. (2021). *Measuring productivity at the industry level: The India KLEMS database*. Reserve Bank of India. <https://rbi.org.in/Scripts/PublicationReportDetails.aspx?UrlPage=&ID=1186>

2. **Filename(s):** \$input_data\\EmpWGbyEDU.dta (file is attached)

(data downloaded (and aggregated) from micro data obtained from government of India, see below)

[the file cannot not be shared on public domain]

- **Source:** MoSPI Microdata Archive: <https://microdata.gov.in>

License terms and access details. The industry-level skill-share dataset was constructed by the authors using unit-level employment data from the National Statistical Office (NSO) *Periodic Labour Force Survey* (PLFS) and earlier National Sample Survey Office (NSSO) *Employment and Unemployment Surveys*, together with the India KLEMS July 2024 database (see above). The PLFS and NSSO unit-level datasets are available upon request through the Ministry of Statistics and Programme Implementation (MoSPI) Microdata Archive (<https://microdata.gov.in/>), subject to MoSPI's data access policies and terms of use. The replication package includes only the derived industry-level skill-share measures and the code used to construct them. It does not redistribute the underlying microdata. Researchers wishing to obtain the unit-level data may do so by requesting access to the PLFS and NSSO unit-level datasets through the MoSPI Microdata Archive (<https://microdata.gov.in/>). After obtaining access to the original data sources, the code included in the replication package can be used to construct the industry-level skill-share dataset and reproduce the analysis. The codes used to aggregate the raw microdata and prepare the essential files for constructing the skill shares are not available, as this work was carried out in collaboration with a colleague from the India KLEMS team.

- **Accessed in:** July 2024
- **Citation:** National Statistical Office (NSO), Ministry of Statistics and Programme Implementation (MoSPI). *Periodic Labour Force Survey (PLFS), Unit Level Data* (various rounds).
https://microdata.gov.in/NADA/index.php/catalog/PLFS/?page=1&sort_order=desc&ps=15&repo=PLFS

THE FOLLOWING DATASETS CONTAIN DATA WITH COPYRIGHT CONCERNS AND CANNOT BE SHARED PUBLICLY:

- **Filename(s):**
 1. "\$input_data\WB_Intan_India_MASTER.dta" (data downloaded from the ProwessDX database, data acquired in April 2023)

[These files contain micro/firm-level data and cannot be shared on public domain]

- **Source:** Centre for Monitoring Indian Economy (CMIE). ProwessDX [database]

License terms and access details. The data were accessed under my institution (University of Groningen)'s licensed IP-based subscription to CMIE's ProwessDX database. Under the CMIE license, authorized institutional users are granted a non-exclusive, non-commercial right to access and use the database for research purposes. The license prohibits redistribution of the underlying data. The replication package includes the code and documentation necessary to reproduce the analysis and does not include the proprietary ProwessDX data. The master files are only for internal use and should not be shared in any public domain. Researchers wishing to obtain the data should request access from the Centre for Monitoring Indian Economy (CMIE). Access is available through an individual subscription or an institutional (IP-based) subscription to ProwessDX. Researchers affiliated with subscribing universities or research institutions may already have access through their institution. Information on obtaining a subscription is available from CMIE.

This data file contains raw firm-level time series data on companies' annual financial statements, sourced from the Prowess database. The dataset provides company identifiers (including firm name, code, and year of incorporation), basic contact information (such as email and website), and a range of industry and sector classification codes. It includes a comprehensive set of annual financial account variables covering income, sales, expenses, assets, and employment. Additionally, the file contains selected variables on ownership structure, sectoral mappings to KLEMS, and industry-level output and investment price deflators.

- **Accessed in:** April 2023

- **Citation:** Centre for Monitoring Indian Economy (CMIE). ProwessDX [database]. Mumbai, India: CMIE. Available at: <https://prowessdx.cmie.com> (login required).

List of Exhibits

Exhibit name	Output filename	Script	Note
Table 2	\$output_table\table_2.csv	\$dofiles\Table2_baseline_Jan2026.do	Found in "\$output_table\\"
Table 3	\$output_table\table_3.csv	"\$dofiles\Table3_GMM_Jan2026.do"	Found in "\$output_table\\"
Table 4	\$output_table\table_4.csv	"\$dofiles\Table4_Add_Controls_Jan2026.do"	Found in "\$output_table\\"
Table 5	\$output_table\table_5.csv	"\$dofiles\Table5_Heterogeneity_Jan2026.do"	Found in "\$output_table\\"
Table 6	\$output_table\table_6.csv	"\$dofiles\Table6_LP_Jan2026.do"	Found in "\$output_table\\"
Table 7	\$output_table\table_7.csv	"\$dofiles\Table7_Emp_Jan2026.do"	Found in "\$output_table\\"
Table 8	\$output_table\table_8.csv	"\$dofiles\Table8_LP_Heter_Jan2026.do"	Found in "\$output_table\\"
Figure 1	"\$output_figure\Figure1A_shr_int.png"; "\$output_figure\Figure1B_shr_int.png"	"\$dofiles\Figure1_intan_shr_Jan2026.do"	Found in "\$output_figure\\"
Figure 2	"\$output_figure\Fig2A_scatter_intan_lab.png"; "\$output_figure\Fig2B_scatter_tan_lab.png"	"\$dofiles\Figure2_scatter_capital_labor_income_Jan2026.do"	Found in "\$output_figure\\"
Figure 3	"\$output_figure\Figure3_Heter.png"	"\$dofiles\Fig3_Heterogeneity_Jan2026.do"	Found in "\$output_figure\\"
Figure A1	"\$output_figure\FigureA1_Kstck_Inv_sector.gph"; "\$output_figure\FigureA1_Kstck_Inv_sector.emf"	"\$dofiles\FigureA1_INV_K_Composition_Jan2026.do"	Found in "\$output_figure\\"
Figure A2	"\$output_figure\FigureA2A_sales_va.emf"; "\$output_figure\FigureA2B_sales_va2.emf"	"\$dofiles\FigureA2_sales_va_Jan2026.do"	Found in "\$output_figure\\"
Figure A3	"\$output_figure\FigureA3_rep_imp_emp_ratio.emf"	"\$dofiles\FigureA3_report_impute_emp_Jan2026.do"	Found in "\$output_figure\\"
Figure A4	"\$output_figure\FigureA4_Kint_sec_Jan2026.emf"	"\$dofiles\FigureA4_dist_intangibles_Jan2026.do"	Found in "\$output_figure\\"
Figure A5	"\$output_figure\ME_lnK_intanX_int_win_size.pdf"; "\$output_figure\ME_lnK_tan_int_win_size.pdf"; "\$output_figure\ME_lnK_intanX_int_win_Firm_age.pdf"; "\$output_figure\ME_lnK_tan_int_win_Firm_age.pdf";	"\$dofiles\FigureA5_Heteorgeneity_firm_age_size_Jan2026.do"	Found in "\$output_figure\\"
Figure A6	"\$output_figure\FigureA6_skill_shr.pdf"	"\$dofiles\FigureA6_SkillShare_Productivity_Jan2026.do"	Found in "\$output_figure\\"
Table A1	\$output_table\Table A1.doc	"\$dofiles\TableA1_sum_stat_by_ind_Jan2026.do"	Found in "\$output_table\\"

Exhibit name	Output filename	Script	Note
			[The columns under “KLEMS” are constructed using data from the India KLEMS database July 2024 version, data accessed around Jan 2025]
Table A2	\$output_table\\TableA2.xlsx	"\$dofiles\\TableA2_DepreciationRates_Jan2026.do"	Found in “\$output_table\\” [The values are taken from Corrado, Hulten and Sichel (2009) and Das et al (2021).]
Table A3	“\$output_table\\Table A3.xls”	"\$dofiles\\TableA3_Capitalstock_by_Asset_Jan2026.do"	Found in “\$output_table\\”
Table A4	“\$output_table\\Table A4.doc”	"\$dofiles\\TableA4_sum_stat_Jan2026.do"	Found in “\$output_table\\”
Table A5	“\$output_table\\table_A5a.csv”; “\$output_table\\table_A5b.csv”; “\$output_table\\table_A5c.csv”; “\$output_table\\table_A5d.csv”	"\$dofiles\\TableA5_robchk_winz_Jan2026.do"	Found in “\$output_table\\”
Table A6	“\$output_table\\table_A6.csv”	"\$dofiles\\TableA6_excl_ext_val_Jan2026.do"	Found in “\$output_table\\”
Table A7	“\$output_table\\table_A7.csv”	"\$dofiles\\TableA7_Emp_Imp_Heter_Jan2026.do"	Found in “\$output_table\\”