



# JOBS DIAGNOSTIC

## ***Jobs Diagnostic Demand Side Tool***

***Authors:*** Reyes Aterido and Adrian Scutaru

***Contact:*** [raterido@worldbank.org](mailto:raterido@worldbank.org)

***Date:*** 16-May-19

## **Guidance for using the Jobs Diagnostic Demand Side Tool**

### Introduction to the tool

The Job Diagnostic Demand Side Tool (JDD) performs a standardized analysis of jobs outcomes of the formal private sector. It uses firm-level data to generate a set of tables, statistics, and graphs to understand private sector profiles, aggregate trends, and firm performances over time. It benchmarks some indicators with countries for which Job Diagnostics exist.

The first section of the analysis looks at the types of firms that compose the private sector (their size, age, ownership, and sectors and locations where they operate). It determines the employment and revenues they generate, the employment and revenues concentration of the sectors where they operate, and how these have changed over time.

The second section focuses on sectoral trends. It determines which sectors employ more workers, what is their labor productivity and their labor costs per worker. It also looks at how many jobs those sectors have created, and changes in their labor productivity and labor cost per worker. Because outputs are generated at different levels of aggregation, we recommend focusing on a predetermined level of aggregation to ensure consistency and correct interpretation.

The last section looks at the firm level. Here outputs refer to the firm. In the previous section, labor productivity in a given sector signifies the total value added generated by the sector divided by its total employment. In this third section, a sector's labor productivity refers to the average of firm level productivity in the sector. The analysis in this section includes job creation, firm productivity, and average wage by type of firm. It determines whether workers are allocated to more productive firms and by means of regression analysis estimates the determinants of these outcomes and their growth rates. For example, regression analysis shows whether productive firms grow faster, or whether younger firms are more productive. The structure of the analysis corresponds to the questions laid out in the accompanying *Guided Enquiry*.

The data is from censuses, administrative data, registries, or representative surveys. Our default for the economic unit is establishment (firm if not available). The tool drops firms that do not have at least one permanent employee and sectors largely not market oriented (such as health, education, and public administration -isic version 4 larger than 84).

## Instructions

User needs to have STATA software, and be proficient in both programming STATA and managing firm level data.

### 1. JDD toolkit includes:

- "WBSPJ\_JDD1\_master.do": runs toolkit
- "WBSPJ\_JDD2\_preparation.do": stores information about the data
- "WBSPJ\_JDD3\_data.do": constructs and cleans dataset
- "WBSPJ\_JDD4\_programs.do": stores programs
- "WBSPJ\_JDD5\_enquiry.do": generates outputs
- "JDD\_benchmark.dta": data of benchmarking indicators
- "JDD\_lifecycle.dta": data of benchmarking indicators for lifecycle of the firm
- "PROD2\_label.dta": industry labels

### 2. Outputs include<sup>1</sup>:

- Figures in \*.gph, \*.png, and \*.eps if required (gph and eps graphs are editable)
- Tables\_JDD\_CCC.xls stores all tables generated by JDD
- Output\_data\_JDD\_CCC.xls stores all data to generate figures
- Regressions\_JDD\_CCC.xls
- TableA1\_CCC.xlsx summary statistics (annex)

Refer to *Enquiry* for list of figures and tables

## Instructions for using JDD<sup>2</sup>:

CREATE three folders: (i) dofiles, (ii) input, (iii) output

PLACE all *dofiles* in "dofiles" folder

PLACE all *datasets* in "input" folder

CLEAN your data (instruction below)

INCLUDE deflators in your data<sup>3</sup> (instruction below)

SAVE **cleaned** dataset "CCC\_raw.dta" and place it in "input" folder

FILL-IN "WBSPJ\_JDD2\_preparation.do" (all "Zz's")

- Section 1: information about the dataset
- Section 2: matches original variables to standard variables

SAVE WBSPJ\_JDD2\_preparation\_CCC.do" (this dofile is unique to the country's database)

RUN TOOL: "WBSPJ\_JDD1\_master.do":

- Stage 1: generates data
  - "WBSPJ\_JDD2\_preparation\_CCC.do" stores globals before sessions begins

---

<sup>1</sup> CCC is three letters WDI country code

<sup>2</sup> Note that JDD will not run if these instructions are not strictly followed

<sup>3</sup> Data is expected to be in nominal values. The toolkit deflates.

- "WBSPJ\_JDD3\_data.do" constructs standardized data and saves it "CCC\_JDDdata.dta"

NOTE1: Once "CCC\_JDD.dta" is saved don't run again

- o Stage 2: generates outputs
  - RUN "WBSPJ\_JDD2\_preparation\_CCC.do" (run always)
  - RUN "WBSPJ\_JDD4\_programs.do" (run always)
  - RUN "WBSPJ\_JDD5\_enquiry.do"

NOTE2: Only WBSPJ\_JDD2\_preparation.do needs user's input (saved as WBSPJ\_JDD2\_preparation\_CCC.do)

NOTE3: once data is saved, only stage 2 is needed

### Instructions for deflators:

SET reference year = 2010

SET deflators = 1 if data is one-year cross-section or is already deflated

SET deflators in units (not percentage)

USE industry level deflators if available

USE WDI GDP deflators if deflators not available

ALL observations MUST have deflators

- o Deflators should match dataset's industry and year (dataset industry code and deflators need to have same standard classification and same number of digits)
- o If there are missing values for deflators in some years or industries, please use your judgement but no observation should lack deflators (i.e. use another industry or an average of similar industries)
- o GENERATE the following variables (**need to be generated and saved in raw data file**):
  1. go\_p: price deflator for gross output (use WDI GDP if not available)
  2. va\_p: price deflator for value added (use WDI GDP if not available)
  3. k\_p: price deflator for capital (use WDI GDP if not available)
  4. mat\_p: price deflator for inputs (use WDI GDP if not available)
  5. wag\_p: price deflator for wages (use cpi if not available)

NOTE4: if any of go\_p, va\_p, k\_p, or mat\_p is available, use it for all others (except wages) rather than WDI GDP deflator.

### Guidelines for basic data cleaning and preparation

#### 1. User: cleaning

- document everything that modifies original dataset
- when cleaning, check always how many firms and what type of firms are dropped
- set data in long format (each row is one firm and year)
- check that firms are economically active (i.e. sales and employment > 0; if sales = 0, and cost of production and wages = 0 consider inactive)

- If number of establishments in commercial agriculture or mining are not representative of the sector, do not keep these sectors
- if formal establishments can be identified, keep only formal establishments
- check that there are no negative values for: employment, age, sales, input costs, wages, capital, etc..
- check abnormal values, outliers and put to missing
- check when firm started operations (it will be used to determine age; this variable would likely need cleaning; age can't be < 0 or very large numbers).
- If age is missing and data is panel, year when firm is first observed (if different than first year of dataset) is considered entry
- check how many observations are missing in key variables and assess whether the variable should not be used
- check whether some variables should be set to 0 when missing (for example missing exports)
- check that establishment's id is unique
- check number of duplicates (id & year should be unique; if everything is the same, assess whether is a duplicate)
- check that id's identify unique firm (i.e. if difference in variables are too large, assess whether are two different firms)
- if panel, check for abnormal or inconsistent jumps in specific variables of one firm (i.e. employment, sales, wage bill). This needs to be solved case by case.

## 2. User: preparation

- match industry code to **UN isicv4** in cleaned dataset (raw\_CCC.dta)
- maximum number of digits used in analysis is 4
- set your industry code variable in 4 digits numbers (isicv4) (if 4 digit industry is not available, have industry variable in biggest isicv4 digits)
- please set location variable in numeric values regions and provinces)
- set region = 1 to be reference region (i.e. most urban, capital city)
- save your data in input folder "CCC\_raw.dta", where CCC is WDI 3 letters country code
- fill-in ALL "Zs" in "WBSPJ\_JDD2\_preparation.do"; save as "WBSPJ\_JDD2\_preparation\_CCC.do" This file stores specific information about your data
- all variables **MUST** be generated (set to missing if not available)
- make sure weights are available if dataset is a survey; set weight = 1 if census
- if panel data but shorter than 5 years, we recommend doing a cross section analysis
- age at entry year is one year old

## 3. Tool: cleaning (stage 1)

- Productivity outliers to missing: (i) 1 percent tails in 2-digit industry distribution (if number of firms in that industry is 20 or more, otherwise 1 percent tails in all firms' distribution); (ii) and extreme values ( $\text{abs}(\log \text{ productivity}) > 16$ ).

- Wages outliers to missing: wages divided by employment is larger/smaller than wage/employment's average plus/minus four times wage/employment's standard deviation of 2-digit industry distribution if number of firms in that industry is 20 or more, otherwise all firms' distribution.
- Materials outliers to missing: if materials divided by sales is larger/smaller than its average plus/minus four times its standard deviation in 2-digit industry distribution (if number of firms in that industry is 20 or more, otherwise in all firms' distribution).
- Capital outliers to missing: if capital divided by sales and capital divided by employment is larger/smaller than its average plus/minus four times its standard deviation in 2-digit industry distribution (if number of firms in that industry is 20 or more, otherwise all firms' distribution).

### Definitions

- Permanent employment is the default (total employment if not available). If duration of employment is available, use "full time equivalent employment".
- Age, size and sector in most outputs are categorical or dummies.
- There are two variables for labor costs: labor cost including benefits is used in aggregate section if available; wag bill is used in firm-level section if available. Firm's average wage is firms wage bill divided by employment.
- Value added is sales minus cost of sales (raw materials and finished goods) minus indirect costs (fuel, electricity, transport, rent, interest, etc...). By residual, value added is profit before tax plus remuneration employees.
- Capital is fixed assets after depreciation and monetary correction if available (default excludes land and buildings).
- Taking logs of sales, output, materials, value added, labor costs and wage bill, and capital:  $\log(X+1e-6)$ .
- Labor productivity is value added per worker if available, otherwise sales or output per worker.

### Construction variables

- Growth is *change[t-1 to t] divided by average [(t-1 + t)/2]* (see Davis and Haltiwanger 1992 and 1999). It is symmetric [-2,2]) monotonically related to the conventional growth rate and a second order approximation of the logarithmic first difference.
- Net job creation:
  - o If data is panel, net job creation is computed at the firm level (job creation at entry is employment at entry; job destruction at exit is minus employment at exit).
  - o If data is a cross section with more than two years, net job creation is the difference in total employment year  $t-1$  to  $t$ .

- Decomposition productivity change (second section - aggregate):

*Productivity change = within + between*

$$\Delta Y_t = \sum_{i=n} \theta_{j,t-k} \Delta y_{i,t} + \sum_{i=n} y_{j,t} \Delta \theta_{j,t}$$

Y is aggregate productivity  
j is sector  
 $\theta_j$  is sector share  
 $y_j$  is sector productivity

- Decomposition productivity change (third section - firm level):

*Aggregate productivity = unweighted productivity + covariance term*

(the covariance term is the labor reallocation effect (how labor shares are allocated to more productive firms)).

$$\Omega_{jt} = \sum_j s_{it} \omega_{it} = \bar{\omega}_{jt} + \sum_j (s_{it} - \bar{s}_{jt}) (\omega_{it} - \bar{\omega}_{jt})$$

$\Omega$  is aggregate productivity  
j is sector  
t is year  
 $s_i$  is firm's share in sector j  
 $\omega_i$  is firm productivity