Data and Replication Information for

"How Do Taxpayers Respond to Public Disclosure and Social Recognition Programs? Evidence from Pakistan"

Joel Slemrod, University of Michigan
Obeid Ur Rehman, University of Michigan
Mazhar Waseem, University of Manchester

April 2020

1. Data

The paper uses two sets of data. First, it uses the public disclosure data for the years 2012-2015. Our definition of the tax year is slightly different from the FBR's. For example, we define the tax year 2012-2013 as 2012, whereas the FBR defines it as 2013. Our data for the tax year 2012 therefore corresponds to the tax year 2013 of the disclosed data and so on. We provide the disclosed data for these four years, separately for parliamentarians and all taxpayers. These data correspond to the tax directories published by the FBR on its website for tax years 2013-2016. Second, we use confidential, administrative data, which were provided by the Federal Board of Revenue (FBR), Pakistan, for the purpose of conducting this research. Any person wishing to obtain access to the data would have to approach the senior management at FBR, making a business case for the access. One of the authors, Mazhar Waseem, however, has an ongoing relationship with the FBR management, and would happily make introductions to researchers who are interested in replicating the results in this paper or for doing any other work they could make a business case for. Similarly, we would be happy to run any sensitivity analyses ourselves and share the results with anyone wishing to run such analyses.

2. Files

The data archive contains the following files. *Slemrod-Rehman-Waseem-2020-Master.do* file details the association between a do file and a figure or table in the paper.

i. README.pdf
ii. Slemrod-Rehman-Waseem-2020-Master.do
iii. Distribution of Names.do
iv. Charts Public Disclosure Response.do
v. Charts Public Disclosure Response – Extensive Margin.do
vi. Charts Status Signalling Response.do
vii. Charts Google Trends.do
viii. Distribution of Names – Original Spellings.do
ix. Charts Public Disclosure Response – SE vs WE.do

x. Distribution of Names -- Baseline.do
xi. Charts Public Disclosure Response – Extensive Margin Robustness.do
xii. Table Public Disclosure Response.do
xiii. Table Public Disclosure Response – Extensive Margin.do
xiv. Table Status Signalling Response.do
xv. Summary Statistics.do
xvi. Table Treatment Control Balance.do
xvii. Table Treatment Control Balance – Balanced Panel.do
xviii. Table Public Disclosure Response – Pre & Post Treatment.do
xix. Table Public Disclosure Response -- Long.do
xx. Table Public Disclosure Response – MPs Placebo.do
xxi. Table Public Disclosure Response – Robustness Name Frequency.do
xxii. Table Public Disclosure Response – With Gender x Year FEs.do
xxiii. Table Public Disclosure Response – With Age x Year FEs.do
xxiv. Table Public Disclosure Response – With Region x Year FEs.do
xxv. Table Public Disclosure Response by Taxable Income.do
xxvi. Table Public Disclosure Response – SE vs WE.do
xxvii. Table Public Disclosure Response -- Heterogeneity.do
xxviii. Table Public Disclosure Response – Extensive Margin Baseline.do
xxix. Table Status Signalling Response -- Placebo.do
xxx. Table Status Signalling Response – By TaxPayer Type.do
xxxi. Table Status Signalling Response -- Heterogeneity.do
xxxii. Table Summary Statistics TPHE.do
xxxiii. Table Earnings Response WE Robustness.do
xxxiv. Name cleaning.do
xxxv. Full Disclosure Clean.do
xxxvi. FullDisclosure2012.dta
xxxvii. FullDisclosure2013.dta
xxxviii. FullDisclosure2014.dta
xxxix. FullDisclosure2015.dta
xl. Parliamentarians2012
xli. Parliamentarians2013
xlii. Parliamentarians2014
xliii. Parliamentarians2015

3. Replication procedure

Define the *project_do*, *project_data*, and *project_output* directories in the *Slemrod-Rehman-Waseem-2020-Master.do* file. Make sure that the data and do files are placed in the defined directories. Run the Slemrod-Rehman-Waseem-2020-Master.do file. It calls the other do files and generates all results in the paper, including those in the Online Appendix. The results in the paper were created using STATA version 14-16.

4. Definition of variables

Definition of all variables used in the project are provided in Appendix A.2 of the paper