The Generalized-Weights Method in Practice: Instructions on the use of R code

Pedro R. D. Bom[*], Heiko Rachinger[†]

March 2024

1 Introduction

This document provides some instructions on the use of the R code 'GW_practice.R', which implements our generalized weights (GW) estimator in meta-analyses with sample overlap. It generalizes the previous Stata code 'GW_stata.do' insofar as (1) it works for meta-analyses based on regression coefficients and on partial correlation coefficients (PCC), (2) it allows for different estimation methods (OLS vs IV), (3) it allows for temporally and spatially aggregated data, and (4) it allows for meta-regressions with moderator variables.

The code requires the user to input the meta-analysis data in an Excel file. The code then reads these data, constructs the variance-covariance matrix resulting from sample overlap (the Ω matrix in the paper), and implements the GW estimator (together with the RE and FE/WLS estimator). The variance-covariance matrix is stored in a Excel file named "'Omega_matrix.xlsx"'. This matrix may be useful to users who wish to visualize the correlation structure or develop variations of the GW estimator.

2 How to use the R Code

To implement the R code 'GW_practice_R' in a meta-analysis setting with sample overlap, please follow these steps:

[*]Deusto Business School, University of Deusto. Address: Hermanos Aguirre Kalea 2, 48014 Bilbao, Spain. Tel.: +34 944 139 290. E-mail: pedro.bom@deusto.es.

[†]Department of Applied Economics, Universitat de les Illes Balears. Address: Cra. de Valldemossa, km 7.5, 07122 Palma de Mallorca, Spain. Tel.: +34 971 179 549. E-mail: heiko.rachinger@uib.es.

1. The code reads the meta-analysis data stored in the first sheet of an Excel file. It assumes the data are stored in columns starting from the second row (the first row is reserved for data headings and can be used at the user's discretion) and must have the following structure:

    - Column A: Primary estimates.
    - Column B: Standard errors of primary estimates.
    - Column C: PCCs in case the meta-analysis is based on PCC, otherwise leave this column should empty.
    - Column D: Standard errors of the PCCs in case the meta-analysis is based on PCCs, otherwise leave this column empty.
    - Column E: Start year of the sample (for simplicity the code assumes that the series starts in the first month/quarter of this year).
    - Column F: End year of the sample (for simplicity the code assumes that the series ends in the last month/quarter of this year).
    - Column G: Estimation method: 0 for OLS, 1 for IV.
    - Column H: Frequency: 'A' for annual data, 'Q' for quarterly data, or 'M' for monthly data (in capitals).
    - Column I: Spatial layer: 'S' for supranational data (highest layer), 'N' for national data (middle layer), or 'R' for regional data (lowest layer) (in capitals).
    - Column J: List of regions: if regional data are used, provide the list of regions included in the dataset (separated only by spaces). If the data are defined at the national or supranational level, leave this column empty. (Note: Make sure the spelling of the regions' names is consistent throughout the file.)
    - Column K: Number of regions: if regional data are used, indicate the total number of regions at this level of aggregation. If national or supranational data, leave this column empty.
    - Column L: List of countries: If regional data are used, provide the list of countries (separated only by spaces) to which the regions belong. If national data are used, provide the list of countries used (separated only by spaces). If supranational data are used, provide the list of all countries (separated only by spaces) the supranational dataset consists of. (Note: Make sure the spelling of the countries' names is consistent throughout the file.)

- Columns M- : Add any number of moderator variables in consecutive columns.

2. Label the Excel file 'data.xlsx' and place it in the working directory specified in R (or change the R working directory to the directory where your file is stored). Alternatively:

   - If you want to give your data file any other name (e.g., 'mymetadata.xls'), you must replace the default name 'data.xlsx' by your new name in line 17 of the code.
   - If your data is saved using the extension .xls, you must replace the default extension .xlsx by the extension .xls in line 17 of the code.

3. For the covariance between OLS and IV estimates, the code uses by default the conservative OLS-OLS formula. If you want to change it to the exact IV-OLS formula, please change line 18 to 'conservativeIV <-0'. (Note: as explained in the paper, the exact formula may, in some applications, cause technical complications by rendering the matrix Ω non-positive definite.)

4. Execute the code.

The code returns five pieces of output:

1. Information about the number of moderator variables: "You work with 'N_mods' moderator".

2. Information about the intended object of the meta-analysis. "You analyze regression coefficients", if columns A and B are full and C is not. "You analyze PCCs" if columns C and D are full "Revise your input: either all have to be regression coefficients, or all PCCs, both together with their SE." if neither columns A and B are full, nor C and D.

3. Information about the OLS-IV formula. If you use the conservative formula, it writes: "You work with the conservative OLS-IV formula".

4. The GW estimate and its standard error (together with the RE and FE/WLS estimator). These results are printed in the main R window. (Note that the code allows for an unrestricted multiplicative constant to the variance-covariance matrix.)

5. The variance-covariance matrix ($\mathbf{\Omega}$ in the paper), from which the weighting matrix $\mathbf{\Omega^{-1/2}}$ is computed. This matrix is stored under the name 'Omega_matrix.xlsx' in the working directory.