Readme - Replication Package for "Estimation of Discrete Games with Weak Assumptions on Information"

Lorenzo Magnolfi*and Camilla Roncoroni†

June 5, 2022

Contents

1 Overview **2**

2 Data Availability Statements **2**

3 Dataset List **2**

4 Computational Requirements **3**

5 Description of Programs **3**

6 Instructions for Replicators **5**

7 List of Exhibits and Programs **6**

*Corresponding Author: Department of Economics, University of Wisconsin-Madison, magnolfi@wisc.edu.
†Department of Economics, University of Warwick, c.roncoroni@warwick.ac.uk.

1 Overview

The replication package contains data and code to generate the figures and tables in the paper. Replication requires around 54 hours for the simulation results (Sections 2-5 of the article), and 309 hours for the empirical application results (Sections 6-7 of the article).

2 Data Availability Statements

The sources of data used in the paper include: (i) data on all supermarkets in Northern and Central Italy from the market research firm IRI, (ii) hand-collected information on shopping malls, obtained from online directories, (iii) data on population from the 2011 Italian Census, (iv) data on average income from the Italian Ministry of the Economy. We refer to these sources, as IRI, Information Resources Srl (2013), ISTAT (2011), and MEF, Italian Ministry of the Economy (2013).

Source (i) is the "Top Trade Guide" dataset, collected and sold by IRI (`https://www.iriworldwide.com/it-it`). While this data may not be disclosed by the authors according to the contract between the authors and the data provider, the data can be purchased by other researchers for the purposes of replication. Source (ii) is collected and made available by the authors. Sources (iii) and (iv) are publicly available for download from ISTAT (`http://dati-censimentopopolazione.istat.it/Index.aspx`) and from the Italian Ministry of the Economy (`https://www1.finanze.gov.it/finanze/analisi_stat/public/index.php?opendata=yes`), respectively.

We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript. The code is licensed under a Creative Commons/CC-BY-NC license.

3 Dataset List

The data files used are as follows:

Data File	Source	Provided
`supermarkets.csv`	(i): IRI, Information Resources Srl (2013)	No
`demographics.csv`	(ii)–(iv): ISTAT (2011) MEF, Italian Ministry of the Economy (2013)	Yes

4 Computational Requirements

Software: The software required for replication includes:

- MATLAB (the code was run with version R2016a), including the additional packages:
  - MPT Toolbox (v. 3.0) - see Kvasnica, Grieder, Baotić, and Morari (2004);
  - CVX (v. 2.1) for MATLAB, including the solver SeDuMi - see Grant and Boyd (2014, 2008);
- AMPL (v. 10) and the solver Knitro (v. 9.1);
- The script Parampl (v. 1.1) - see Olszak and Karbowski (2018);
- Stata 15.1.

Memory and Runtime: The simulation results (Sections 2-5 of the paper) and some of the empirical application results (Sections 6-7 of the paper) are obtained with a 2014 Windows 10 desktop machine with 4 cores and 32GB of RAM (*PC*). The total runtime on *PC* is approximately 55 hours. Other empirical application results (Sections 6-7 of the paper) are obtained on a Linux server with 36 cores and 768GB of RAM (*server*). The total runtime on *server* is approximately 308 hours.

5 Description of Programs

- The MATLAB script `Table1.m` generates the Excel file `Table1.xls`, which contains the results in Table 1 of the paper. The main script uses auxiliary functions `GenData_Unif.m`, `allcomb.m`, and `computeBCE_par_Unif`. `allcomb.m` is written by van der Geest (2013). The function `computeBCE_par_Unif.m` relies on the function `fprintAmplParamCLSU.m` to produce data files for AMPL, as well as on the AMPL scripts `Dual_Problem_Par.mod` and `Dual_Problem_Parampl.run`. The function `fprintAmplParamCLSU.m`, used here as well as in many other steps in this package, is included in Su and Judd (2012).
- The MATLAB script `Table2.m` generates the three panels in Table 2 of the paper. The script relies on a number of functions:
  - `GenData.m` generates the simulated data needed to perform the identification exercise in the table;
  - `computeBCE_par.m` and `computeBCE_par_corr.m`, respectively, generate identified sets for panels A,B and C;

  - The function `computeBCE_par.m` uses the AMPL scripts `Dual_Problem_Parampl.run` and `Dual_Problem_Par.mod` to solve the optimization step required to find the identified set;
  - Similarly, the function `computeBCE_par_corr.m` uses the AMPL scripts `Dual_Problem_Parampl_corr.run` and `Dual_Problem_Parampl_corr.mod` to solve the optimization step required to find the identified set.

- The MATLAB script `Figure2.m` generates the two panels of Figure 2 in the paper, using the two AMPL scripts `Polytope.mod` and `Polytope.run`.

- The MATLAB script `Figure3.m` generates the three panels of Figure 3 in the paper. The script relies on a number of functions:

  - `GenData.m` and `GenData_PublicSignal.m` generate the data;
  - `computeMXNE.m`, `computeR1.m`, `computeR2.m`, `computeBCE_par.m`, `computeBCE_PublicInfo.m`, `computeBCE_PublicInfo_higherBase.m` compute identified sets for different assumptions on information and equilibrium;
  - The function `computeBCE_par.m` calls AMPL and Knitro. AMPL uses the files `Dual_Problem_Par.mod` and `Dual_Problem_Parampl.run`.

- The MATLAB script `BCE_estim_CF.m` produces estimation results in Table 3, Column I (lines 1-502), and counterfactual results in Table 4, Column I, and Table 5 (lines 503-745). The script relies on a number of functions:

  - The MATLAB functions `loadData.m`, `prepdata.m` and `QuantTrue.m` prepare the data for estimation;
  - MATLAB functions `BCE_Inf_CC_cluster.m` and `BCE_Inf_CC_cluster_m.m`, and AMPL scripts `Dual_Problem_m.run` and `Dual_Problem.mod` are used in the estimation step;
  - The function `anneal.m` is written by Vandekerckhove (2006).
  - MATLAB functions `BCE_VLI_CF.m`, `BCE_Inf_forCF_server.m`, `BCE_FLI_CF.m` and AMPL scripts `CF_OneX.run`, `CF_REps_OneX_w9_2.mod`, `Dual_Problem_forCF.run`, `NewCF_OneX_1_knitro.run`, and `NewCF_newdef.mod` are used in the counterfactual step;
  - The script produces the data files `VLI_Interval_bef.mat` and `VLI_Interval_aft.mat`.

- The MATLAB script `NE_estim_CF.m` produces estimation results in Table 3, Column II (lines 1-428), and counterfactual results in Table 4, Column II (429-672), and Figure 4 (673-772). The script relies on a number of functions:

  - MATLAB functions `Nash_Inf_CC_cluster.m` and `Nash_Inf_CC_cluster_m.m` are used in the estimation step; these functions rely on `CorrN_Quadra.m` to compute event probabilities.
  - The MATLAB function `Nash_CFcomp.m` is used in the counterfactual step;

- The MATLAB script `BNE_estim_CF.m` produces estimation results in Table 3, Column III and counterfactual results in Table 4, Column III. Part 1 of the script (lines 1-256) performs the estimation step. Part 2 of the script (lines 257-438) performs countefactual policy exercises. The script relies on a number of functions:

  - AMPL scripts `Data_3pl.run`, `3pl.mod`, and `DataCF_3pl.run`, `3pl_CF.mod` execute the maximization step necessary for estimation and counterfactual prediction, respectively.

- The Stata script `RedForm_CF.do` produces the reduced form counterfactual results in column IV of Table 4.

6 Instructions for Replicators

- On *PC*, unzip the package in a `\home` folder; the AMPL executable `ampl.exe` should be installed in a folder `\home\ampl`. Create an empty `\home \Output` folder for temporary files. All MATLAB packages listed above, and Parampl, should be installed. Tables 1-2 and Figures 2-3 can be obtained by running the corresponding MATLAB scripts e.g., run `Table1.m` to obtain Table 1. The steps to follow are:

  1. Run `Figure2.m` (2 minutes on *PC*)
  2. Run `Table1.m` (6 minutes on *PC*)
  3. Run `Figure3.m` (6.5 hours on *PC*)
  4. Run `Table2.m` (47 hours on *PC*)

- Tables 3-5 and Figure 4 are obtained as follows:

  1. Run `BCE_estim_CF.m` (162 hours on *server*)
  2. Run `NE_estim_CF.m` (147 hours on *server*)
  3. Run `BNE_estim_CF.m` (1.2 hours on *PC*)
  4. Run `RedForm_CF.do` (2 minutes on *PC*)

  These steps require the two data files `supermarkets.csv` and `demographics.csv`.

7 List of Exhibits and Programs

The package reproduces the tables and figures in the paper. Excel files corresponding to the tables, and eps files for the figures, are also included in the folder `Figures&Tables`.

Exhibit	Programs	Output File	Note
Table 1	`Table1.m`	`Table1.xls`	
Table 2	`Table2.m`	`Table2.xls`	
Table 3			
Col. I	`BCE_estim _CF.m` (lines 1-502)	`Table3_colI_II.xls`	
Col. II	`NE_estim _CF.m` (lines 1-428)	`Table3_colI_II.xls`	
Col. III	`BNE_estim _CF.m` (lines 1-256)	`Table3_colIII.xls`	
Table 4			
Col. I	`BCE_estim _CF.m` (lines 503-745)	`Table4_colI_II.xls`	
Col. II	`NE_estim _CF.m` (lines 429-672)	`Table4_colI_II.xls`	
Col. III	`BNE_estim _CF.m` (lines 257-438)	`Table4_colIII_IV.xls`	
Col. IV	`RedFormCF.do`	`Table4_colIII_IV.xls`	
Table 5	`BCE_estim _CF.m` (lines 53-745)	`Table5.xls`	
Figure 1			Schematic similar to earlier literature e.g., Tamer (2003)
Panel A		`Fig1_A.eps`	
Panel B		`Fig1_B.eps`	
Panel C		`Fig1_C.eps`	
Figure 2	`Figure2.m`		
Panel A		`Fig2_A.eps`	
Panel B		`Fig2_B.eps`	
Figure 3	`Figure3.m`		
Panel A		`Fig3_PanelA.eps`	
Panel B		`Fig3_PanelB.eps`	
Panel C		`Fig3_PanelC.eps`	
Figure 4	`NE_estim _CF.m` (lines 673-772)		
Panel A		`Fig4_A.eps`	
Panel B		`Fig4_B.eps`	

References

Grant, M. and S. Boyd (2008): "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ed. by V. Blondel, S. Boyd, and H. Kimura, Springer-Verlag Limited, Lecture Notes in Control and Information Sciences, 95–110, http://stanford.edu/~boyd/graph_dcp.html.

——— (2014): "CVX: Matlab Software for Disciplined Convex Programming, version 2.1," http://cvxr.com/cvx.

IRI, Information Resources Srl (2013): "Top Trade Guide," https://www.iriworldwide.com/it-it.

ISTAT (2011): "2011 Population and housing census," http://dati-censimentopopolazione.istat.it/Index.aspx, accessed in May 2022.

Kvasnica, M., P. Grieder, M. Baotić, and M. Morari (2004): "Multi-parametric toolbox (MPT)," in *International workshop on hybrid systems: Computation and control*, Springer, 448–462.

MEF, Italian Ministry of the Economy (2013): "2013 Income Tax (IRPEF) Data," https://www1.finanze.gov.it/finanze/analisi_stat/public/index.php?opendata=yes, accessed in May 2022.

Olszak, A. and A. Karbowski (2018): "Parampl: A Simple Tool for Parallel and Distributed Execution of AMPL Programs," *IEEE Access*, 6, 49282–49291.

Su, C.-L. and K. L. Judd (2012): "Constrained optimization approaches to estimation of structural models," *Econometrica*, 80, 2213–2230.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

van der Geest, J. (2013): "allcomb," https://www.mathworks.com/matlabcentral/fileexchange/10064-allcomb-varargin, MATLAB Central File Exchange.

Vandekerckhove, J. (2006): "anneal," https://www.mathworks.com/matlabcentral/fileexchange/10548-general-simulated-annealing-algorithm, MATLAB Central File Exchange.