# DATA ARCHIVE

## Asset Integration and Attitudes to Risk: Theory and Evidence

by

Steffen Andersen, James C. Cox, Glenn W. Harrison,

Morten I. Lau, E. Elisabet Rutström and Vjollca Sadiraj

November 2017

As explained in the paper, a central feature of our data analysis is to merge data from standard experiments with Danish subjects and confidential data from the Danish Registry on their wealth. The former data are public, in the usual manner. The latter data may only be accessed by accredited researchers, in the manner explained in the paper and in detail in Appendix E of the Online Appendices. We attach that Appendix E at the end of this document, for convenience.

Hence the data archive consists of a mix of "data and code" for the experimental data and analyses, along with "code only" for the analyses undertaken with the merged data.

However, as an interim step we provide "code and simulated data" to replicate the logic of the estimation undertaken with the merged, confidential data. The simulated data are generated by obtaining matrices of means and correlations from the correct Registry data, exporting these matrices, and then "reconstituting" the data from these matrices, just as one might reconstitute dried mushrooms (http://www.recipetips.com/glossary-term/t--33989/reconstitute.asp). These matrices were exported from the server with the permission of *Statistics Denmark*, of course, and provide a general, albeit second-best, way to replicate estimates from a confidential database. The simulated data mimic in expectation the confidential data, most critically on net wealth, and we then merge those simulated data with the actual data from the respondent's experimental responses. We can then undertake the same estimations outside the server as the ones we undertake inside the server with the actual data. We did this as a way of debugging code without being on the server, and as a way of documenting procedures outside the server.

All files ending in ".do" are *Stata* command files. All files ending in ".dta" are *Stata* data files. All files ending in ".nb" are *Mathematica* notebooks. All files ending in ".wms" are *Maple* command files. Version 13.0 of *Stata*, version 11.0 of *Mathematica*, and *Maple 2016* was used for all calculations. Often we ran the *Stata* code on version 15.0 of *Stata*, but using "version control" to emulate version 13.0. The operating system for off-server estimations was *Windows 10 Pro*. The operating system for on-server estimations was *Windows Server*.

The data archive contains the following directories.

1. **AI** contains all analysis files, specifically:
    a. This file, **Asset Integration and Attitudes to Risk – Theory and Evidence DATA ARCHIVE.pdf**, as a **README.pdf** file.
    b. **AI.do** runs all *Stata* files under *Stata* version 13.0 (using "version control" when later versions of *Stata* are employed).
    c. **Data.do** generates the data using data in the DATA directory. The initial code runs off-server using simulated data from estimated correlation matrices from *Statistics Denmark*. The marked-out end code documents how the data is collated on the server.
    d. **MLfunctions.do** contains all likelihood functions used.
    e. **Analysis_A.do** contains code to re-estimate all models and summary statistics on representative models using simulated data. The current code will estimate models off-server. Log files of models run on the server are placed in the ServerLogs directory.
    f. **Analysis_I.do** analyses the individual data: due to the confidentiality of merged

Registry data for individuals no log files on individual estimates may be provided.

g. **AI figure 2 - black and white.nb** generates Figure 2. For a glorious color version, use **AI figure 2.nb**.

h. **AI figure B1.nb** generates Figure B1.

i. **AI figure B2.nb** generates Figure B2.

j. **AI_CodeOnly_17.wms** generates the CE and Ratio values for the numerical examples in Panel B of Table 2 and 3, and both panels of Tables C1 and C2. The file **AI_CodeOnly_17.pdf** displays the *Maple* command file, and the spreadsheet **AI_17.xls** reports the numerical results used in these Tables. This code also generates the "edge cases" reported at the end of §4.C and §4.E, as well as the claims immediately prior that "one can verify that" certain calibration puzzle patterns will not be observed with our estimates.

2. **DATA** contains all data files, specifically

   a. **ExperimentalData.dta**: our off-server experimental data.

   b. **meanM_E.dta**: the mean of variables used to simulate off-server demographics.

   c. **corrM_E.dta**: the correlation matrix of variables used to simulate off-server demographics.

3. **ServerLogs** contains log files of models estimated on the server, some of which may include confidential data (the directory **ToMove** contains the log files that can be moved off the server). This directory is used to write log files on the server that must be examined by "hand" to determine if they contain confidential information. It may also contain some files run off the server with simulated data (and these files are provided with the distribution archive).

4. **ToMove** contains results from running these programs on the server and off-server, with the qualification noted in item 1(f) above. This directory is used to write log files that we know *a priori* do not need to be examined by "hand" to determine if they contain confidential information, as well as log files from **ServerLogs** (when on the server) that have been cleared for distribution off the server.

5. **TempData** contains temporary data files generated by *Stata* and *Mathematica*.

The data set generated on the server contains the following variables:

```
*********************************
* UNIQUE IDENTIFIER
*********************************
pnr

*********************************
* Experimental Variables
*********************************

sampleIDX   "Sample identifier  1, 2 and 3"
decision    "Decision in task 0 or 1"
task        "Task number - 1 through 60"
VPRIZEA1    "Option A: Amount of the First Prize"
VPRIZEA2    "Option A: Amount of the Second Prize"
VPRIZEA3    "Option A: Amount of the Third Prize"
```

```
VPRIZEA4    "Option A: Amount of the Fourth Prize"
VPRIZEB1    "Option B: Amount of the First Prize"
VPRIZEB2    "Option B: Amount of the Second Prize"
VPRIZEB3    "Option B: Amount of the Third Prize"
VPRIZEB4    "Option B: Amount of the Fourth Prize"
probA1      "Option A: Probability of the First Prize"
probA2      "Option A: Probability of the Second Prize"
probA3      "Option A: Probability of the Third Prize"
probA4      "Option A: Probability of the Fourth Prize"
probB1      "Option B: Probability of the First Prize"
probB2      "Option B: Probability of the Second Prize"
probB3      "Option B: Probability of the Third Prize"
probB4      "Option B: Probability of the Fourth Prize"

*****************************************
* Demographics from Social Security Register
*****************************************

age         "Age in years"
married     "Married"
male        "Male"
hsize       "Household size in count"
kom         "Municipality of Household"

************************************************************
* Wealth information from Tax Register - All amounts in DKK
************************************************************

income            "Total income"

assets_total      "Total assets"
assets_realestate "Real estate"
assets_stocks     "Shares and mutual funds"
assets_banks      "Deposits in banks"
assets_bond       "Bonds and mortgages"
assets_pension    "Pension"
assets_cars       "Cars"

liab_total        "Total liabilities"
liab_bankdebt     "Non-mortgage debt in financial institution"
liab_mortgage     "Mortgage"
liab_private      "Privately issued debt"

wealth_net        "Net wealth"
wealth_T          "Net wealth truncated at exp(1) = 2.78 kroner"
truncated         "Indicator for wealth being truncated"
```

For numerical reasons we truncate net wealth at exp(1) ≈ 2.78 kroner ≈ USD 0.5, and refer to this as "zero" in the paper.

## Appendix E: Data Access (FROM ONLINE APPENDICES)

All computer code, and data from the laboratory experiments, is available on request. It may also be downloaded from http://cear.gsu.edu/gwh/. All estimations were undertaken with version 13 of *Stata* (or version 15, under "version control" to mimic version 13).

Access to Danish micro data ("the Registry") follows the Act on Processing of Personal Data (in Danish, the *Lov om Behandling af Personoplysninger*). This requires a notification to the Danish Data Protection Agency whenever data are made available to researchers. Access can only be granted to researchers in authorized environments. Authorizations can be granted to public research and analysts environments (e.g., in universities, research institutes, and ministries) and to research organizations as a part of a charitable organization. Certain groups in the private sector can also obtain authorization. Only Danish institutions are granted authorization. Foreign researchers can have access to Danish micro data if they are affiliated with an appropriate Danish institution. Visiting researchers can have remote access from a workplace in the Danish research institution during their stay in Denmark, and under the Danish authorization.

Generally, data for a new Registry project is obtained by the following steps:

- an appropriate Danish institution grants permission for the research project;
- the Department of Data for Research at *Statistics Denmark* is contacted at http://www.dst.dk/en/TilSalg/Forskningsservice;
- a Project description is delivered to Statistics Denmark for approval;
- data costs are calculated and a contract is signed; and
- data is extracted to the project.

The procedure is described in detail at:

http://www.dst.dk/-/media/Kontorer/13-Forskning-og-Metode/Step-by-step-procedures-for-researchers-access-to-Microdata.pdf?la=en

In the Online Appendices we append this document.

The additional experimental data, which we have collected ourselves, is subsequently obtained by being transferred from our project at *Statistics Denmark* to the new project.

In practice, another way to gain access is for researchers to contact a researcher at an appropriate Danish institution that already has an approved project and is willing to provide access. That researcher would then facilitate contact with *Statistics Denmark*, and access would only be allowed if the new project was subsequently approved. So the only way that this path differs is that the "first contact" is through an existing researcher, rather than *Statistics Denmark*. In the end, direct approval from *Statistics Denmark* is always needed.