README file

Data archive: Hilber, Christian A.L. and Tracy M. Turner, "The Mortgage Interest Deduction and its Impact on Homeownership Decisions," *Review of Economics and Statistics* (2014).

We following REStat suggested structure as proposed by McCullough's article in Econ Journal Watch (Volume 4, Number 3, September 2007, pp 326-337):

[A] The README file should list all the replication files with a brief description of each. It should clearly indicate which programs correspond to what results in the paper.

(1) The data preparation programs are written in UNIX SAS, version SAS (r) 9.1 (TS1M3); comments and guidance are given in each program with the code. Each program used is included. The programs read in and code up the original data (from multiple sources) to generate the file dataset "pan7807v17.csv."

In addition to each SAS program, we provide a summary table that lists all of the SAS programs used in sequence, their purpose, their input data and their resulting output data. This summary table is in the archive document, "README_Data preparation SAS files description.pdf."

(2) All analysis is conducted in STATA version SE 13.1 and replicated by the authors using STATA version SE 13.1. The STATA "do" file is included in the archive, which provides the exact code to replicate the results, explanation on which code corresponds to which results in the paper, and our step by step commentary on the code, and the STATA log file.

[B] The README file should also contain a data dictionary that defines each variable and gives the provenance of all the data.

We include "README_data dictionary.pdf"

[C] The README file should identify the version of the software used (by version number and/or release date) and similarly for the operating system on which the software runs.

The SAS programs are run in version SAS 9.1.3 Service Pack 4 and executed in the UNIX environment on the SunOS 5.10 platform. The STATA programs are desktop programs run using STATA version SE 13.1.

[D] All data should be provided in ASCII format, and the version of the code submitted to the archive should call these same ASCII files.

The data files are in ASCII format. All original data are provided, except for MATFAM00.txt, which is confidential GECODE data, also known as "Geospatial" data, available only through special contract with the PSID. We use the confidential data file in our analysis to merge census tract- and MSA-level data to the public use PSID

household data that are the core of our analysis. For complete instruction on how to obtain the confidential GEOCODE data, see http://simba.isr.umich.edu/restricted/Geospatial.aspx

[E] Authors should provide the original data from which the final dataset is derived and all instructions/code necessary to turn the original data into the dataset analyzed.

As we state above in part [D], we provide the public-use data.

[F] The author should provide code such that the data and code, when placed in the same subdirectory, will execute. Also, the output from doing so should be provided. The author should check to make sure that this runs correctly and produces the results in her paper.

Kindly find this information detailed above in part [A].

[G] The archive should list each paper regardless of whether the paper has been exempted from the rules. Sometimes a paper has been exempted from the data requirement. In such cases, the archive should say that the paper has been exempted from the data requirement, and the code should still be required.

We fully comply and provide all code and original data files that are "public use" data. As we state above in part [D], the confidential GEOCODE data, MATFAM00.txt, are not provided by us. Complete information on how to obtain these data is available at:

see http://simba.isr.umich.edu/restricted/Geospatial.aspx