

This archive includes the raw data and computer code required to reproduce the empirical illustration and Monte Carlo results presented in the paper “Sparse network asymptotics for logistic regression under possible misspecification”.

Data Availability Statement

The empirical illustration in the paper uses information on firm-bank relationships from drawn from the *Dealscan* dataset. The exact extract used in the paper corresponds to syndicated loan deals made in the first six months of 2003 as collected by Jiawei Chen and Kejun Song (2013) for their paper “Two-sided matching in the loan market” in the *International Journal of Industrial Organization*. I am grateful to Jiawei for kindly sharing his estimation sample with me. I used his [200306.xls](#) file for the results presented in Table 1 of the paper; this file is included in this data archive. Researchers interested in the balance of this dataset should contact Jiawei. I have also included their “[Explanation of variables.xlsx](#)” file. This file gives brief variable definitions. Users with specification questions about variable construction should refer to the Chen and Song (2013) paper.

Both of these files are also included in plain text format.

Code Execution Instructions and Computation Requirements

Both the empirical illustration and Monte Carlo experiments are implemented in Python 3.9 Jupyter Notebooks. There are two Notebooks in this repository, one for the empirical illustration and one for the Monte Carlo experiments.

1. `Sparse_Network_Asymptotics_Notebook_Empirical_Illustration.ipynb`
2. `Sparse_Network_Asymptotics_Notebook_Monte_Carlo.ipynb`

These Notebooks make use several Python packages; most of these are standard (e.g., pandas, numpy), but a few are not (e.g., geopy). The required packages are listed and loaded in the first code block in the two Notebooks. To replicate the results users will need to have these packages available in your Python environment. If you are not familiar with how to maintain a Python environment I recommend using Anaconda (<https://www.anaconda.com/>).

The paper also utilizes the *ipt* and *netrics* packages. These are codebases that I have created myself. These codebases are available on GitHub at

<https://github.com/bryangraham/ipt>

and

<https://github.com/bryangraham/netrics>

There are several ways to import these packages into Python. Perhaps the easiest approach is the one taken in the included notebooks. Copy the “ipt” and “netrics” folders in each of the packages (including all subfolders) to your local system. Subsequently load them as follows:

```
import sys
sys.path.append('/Users/bgraham/Dropbox/Sites/software/ipt/')
sys.path.append('/Users/bgraham/Dropbox/Sites/software/netrics/')

# Load ipt and netrics modules
import ipt as ipt
import netrics as netrics
```

You will, of course, need to modify the directories above to the appropriate location on your system.

The markdown boxes in the Notebooks include additional information which may aid replication on your machine. For example, there are links to extended blog posts about both *ipt* and *netrics* packages.

Although I did use a fixed random seed, the Monte Carlo results may not numerically replicate exactly depending on the peculiarities of your machine and Python instance. However, they should replicate up to simulation error.

The author executed these Notebooks on a 2017 MacBook Air (1.4 GHz Dual-Core Intel Core i7). Running the first Notebook should take no more than a few minutes, the second may take as much as several hours (although my machine is quite old and slow).

Finally, Python goes through frequent updates. It is quite likely that this code will not execute properly under future updates of Python and/or its associated packages. In general, it is quite easy to “patch up” code to run on new releases, but users not conversant in Python may find it useful to solicit help from a colleague who should this situation arise.

List of Tables and Figures

Tables 1 and 2 in the paper are reproduced by the “Empirical_Illustration” Notebook.

Tables 3 and 4 (reported in Appendix C) are reproduced by the “Monte Carlo” Notebook.

Exact mappings from code output to the tables is described in the markdown boxes in the Notebooks as well as the code output itself.

Additional Comments

I appreciate your interest in my work. Please feel free to reach out with any questions or concerns at bgraham@econ.berkeley.edu. I am not always able to respond to requests in a timely manner, but it is my goal to do so! Feel very free to adapt and modify the code for your

own purposes. Instructions for how to cite the codebases are included at the top of the two Notebooks.

Finally, if you find any errors, please let me know.

-Bryan

References

Chen, Jiawei and Song, Kejun. (2013). "Two-sided matching in the loan market" *International Journal of Industrial Organization* 31 (2): 145 – 152.