**Replication data and codes**

"When to Be Discrete: The Importance of Time Formulation in the Modeling of Extreme Events in Finance" by Katarzyna Bien-Barkowska and Rodrigo Herrera

Data of the reproducibility package: 7 April 2025.

Authors:

Katarzyna Bień-Barkowska, Institute of Econometrics, Warsaw School of Economics, e-mail: [katarzyna.bien@sgh.waw.pl](mailto:katarzyna.bien@sgh.waw.pl)

Rodrigo Herrera, University of Talca, Faculty of Economics and Business, e-mail: rodrigo.herrera@utalca.cl

Structure of the reproducibility package:

In the folder ***data*** there is the raw (input) data containing the daily log returns both in the MS Excel format (.xls) and the APTECH Gauss format (.dat):

Dow Jones 30 index : DowJones.xlsx or DowJones.dat

S&P 500 index : SP500.xlsx or SP500.dat

NASDAQ Composite index: NASDAQ.xlsx or NASDAQ.dat

Wilshire 5000 index: Wilshire.xlsx or Wilshire.dat.

The dates are in column A (from 2 Jan 1981 to 30 Dec 2022), and the corresponding daily log returns are in column N (of the MS Excel worksheet).

The folder ***data*** includes also the initial parameter values which are necessary for the Maximum Likelihood estimation procedures encoded in the APTECH Gauss programmes.

The folder ***output\_data*** includes necessary intermediate datasets.

The folder ***replicated\_results*** includes the replication results (tables and figures) which can be rewritten by running the attached programming codes.

The only exception is the folder 'Table 1', which contains the estimates of the extremal index reported in Table 1.

The programming software used:

* Matlab R2024b (along with the Statistics and Machine Learning Toolbox). Authors have a total-headcount academic licence available at the Warsaw School of Economics. The licence number is [40950635](https://ch.mathworks.com/licensecenter/licenses/40950635/7889763) (user name: Katarzyna Bień-Barkowska, Warsaw School of Economics).
* APTECH Gauss 24, along with the Maxlik and CML modules. The licence number is 16916856989 (user name: Katarzyna Bień-Barkowska, Warsaw School of Economics).
* R 4.4.3 which is a free software environment for statistic computing. The R packages to be installed are this.path; rugarch; esback; evd,readxl (versions available as on 1 April 2025).

The following programming codes can be used to replicate the results in the article:

**Figure\_1\_2.m** replicates Figure 1 and Figure 2 in the manuscript. This Matlab script uses the datasets with the daily log returns imported to Matlab from the ***data*** folder (i.e., DowJones.xlsx, NASDAQ.xlsx, SP500.xlsx, and Wilshire.xlsx). By running this code the figures are saved as pdf files in the **replicated\_results** folder.

**Figure\_3.m and Figure\_4.m** replicate Figure 3 and Figure 4 in the article. These Matlab scripts do not use any external data. The figures are saved as pdf files in the **replicated\_results** folder.

**Figure\_5.m** (or **Figure\_D12.m**) replicates Figure 5 (or Figure D.12) in the article. By running this code the figure components are saved as pdf files in the **replicated\_results** folder. **Figure\_5.m** (or **Figure\_D12.m**) code uses intermediate datasets containing the Monte Carlo simulation results (available in the folder **output\_data/Simulation).** These intermediate datasets can be replicated by running the APTECH Gauss Script **Monte\_Carlo\_Simulation.gss**.

**Figure\_6\_and\_9.m** replicates Figure 6 and Figure 9 in the manuscript. This Matlab script uses the intermediate datasets with the Hoga scoring functions (corresponding to different threshold u values) . These intermediate datasets are available in the folder **output\_data** (*Scores\_rolling\_GAS\_type\_explanatory* (...).dat ), and can be replicated by running the APTECH GAUSS code: **Estimation\_of\_SPOT\_models\_for\_different\_u\_levels.gss.**

**Figure\_7\_8.m** replicates Figure 7 and Figure 8. This Matlab code uses intermediate datasets containing the trajectory of conditional expectations, hazard functions, VaR and ES estimates, etc., associated with the inter exceedance times used in SPOT models (for the optimal threshold u level). These intermediate datasets are available in the folder **output\_data** (*VaR* (...).dat), and can be replicated using the APTECH GAUSS code: **Estimation\_of\_SPOT\_models\_for\_best\_u\_level.gss** (by setting the option **printout=1** in the opening part of the code, as the default value is **printout=0).**

**Estimation\_of\_SPOT\_models\_for\_best\_u\_level.gss** replicates the Table 3 in the manuscript and Table G6 (in Appendix). This APTECH Gauss code uses the datasets on daily log returns imported from the ***data*** folder (DowJones.dat, NASDAQ.dat, SP500.dat, and Wilshire.dat). The replication of the results is available for **printout=0** (which is the default value). By running this code the corresponding replication results will also be automatically printed to the **replicated\_results** folder as the text file: **Replicated\_Table3\_and\_Table\_G6.txt**

**Estimation\_of\_GARCH\_models\_and\_inference.R** can be used to estimate the GARCH models (utilizing the rugarch library). This R programming script delivers the intermediate datasets with the avarage FZ0 scoring functions in the folder **output\_data**, *FZ0\_Loss*...txt). By running this R script the results from Table 4, Table 5, and Table G7 will aslo be replicated and printed to **replicated\_results** folder as **Replicated\_Table\_4.txt**; **Replicated\_Table\_5.txt**; and **Replicated\_Table\_G7.txt**

**Figure\_10.m** replicates Figure 10 in the manuscript. This Matlab m-file uses the intermediate datasets generated by **Estimation\_of\_GARCH\_models\_and\_inference.R** script. The figure is saved as pdf files in the **replicated\_results** folder.

**Figure\_C11.m** replicates Figure C11 in the manuscript. This Matlab script does not use any external data.

The individual files can be executed at any order to replicate the results.

Computation times:

The results were replicated on PC Intel(R) Core(TM) i7-14700KF 3.40 GHz; 32,0 GB RAM; Windows 11 Home

No parallel computing was used. The running times were:

**Figure\_1\_2.m**: 1.45 seconds;

**Figure\_3.m:** 7.68 seconds;

**Figure\_4.m:** 1.30 seconds;

**Figure\_5.m:** 1.49 seconds;

**Figure\_6\_and\_9.m:** 4.10 seconds;

**Figure\_7\_8.m:** 4.87 seconds;

**Figure\_10.m:** 2.21 seconds;

**Figure\_C11.m:** 4.24 seconds;

**Figure\_D12.m:** 1.14 seconds;

**Estimation\_of\_SPOT\_models\_for\_best\_u\_level.gss:** 4 minutes 45 seconds;

**Estimation\_of\_SPOT\_models\_for\_different\_u\_levels.gss:** 2 hours 44 minutes 28 seconds (this file generates the intetmediate datasets which are already available in the folder **output\_data**);

**Estimation\_of\_GARCH\_models\_and\_inference:** 15 minutes.