This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

# Part 1: Data

☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

## Abstract

Two external datasets are used. (1) Daily returns on selected stocks, the Market portfolio, and factors of Fama and French from 1993-01-05 to 2009-01-30, used for pairwise and joint independence testing among industry sectors. (2) TCGA-SARC bulk RNA-seq gene expression data (STAR–Counts, primary tumors, open access) for 259 soft-tissue sarcoma samples, used for detecting higher-order dependence among cytotoxic immune, ECM/stromal, and proliferation gene modules.

## Availability

☒ Data **are** publicly available.
☐ Data **cannot be made** publicly available.

### Publicly available data

☐ Data are available online at:

☐ Data are available as part of the paper's supplementary material.

☐ Data are publicly available by request, following the process described here:

☒ Data are or will be made available through some other mechanism, described here:

**Dataset 1 (Finance):** The Finance data can be accessed by the following R code:

```
install.packages("gmm")
library(gmm)
data <- data(Finance)
```

**Dataset 2 (TCGA-SARC):** The TCGA-SARC RNA-seq gene expression data are publicly available from the Genomic Data Commons (GDC) and can be downloaded using the `TCGAbiolinks` R/Bioconductor package. The download and preprocessing script is provided at `real-data/gene-triplet/01_download_data.R`. Alternatively, the preprocessed gene Z-scores used in the analysis are included in the repository as `real-data/gene-triplet/TCGA_SARC_gene_zscores.rds`.

## Description

### File format(s)

☐ CSV or other plain text.
☒ Software-specific binary format (.Rda, Python pickle, etc.): .Rda, .rds
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (please specify):

**Data dictionary**

- ☐ Provided by authors in the following file(s):
- ☐ Data file(s) is(are) self-describing (e.g., netCDF files)
- ☒ Available at the following URLs:
  - Finance: https://www.rdocumentation.org/packages/gmm/versions/1.8/topics/Finance
  - TCGA-SARC: https://portal.gdc.cancer.gov/projects/TCGA-SARC (accessed via `TCGAbiolinks`)

# Part 2: Code

## Abstract

The simulation code and real data analysis code are provided. The detailed instruction can be found in the README file in the GitHub provided below.

## Description

**Code format(s)**

- ☒ Script files
  - ☒ R
  - ☐ Python
  - ☐ Matlab
  - ☐ Other:
- ☐ Package
  - ☐ R
  - ☐ Python
  - ☐ MATLAB toolbox
  - ☐ Other:
- ☐ Reproducible report
  - ☐ R Markdown
  - ☐ Jupyter notebook
  - ☐ Other:
- ☐ Shell script
- ☐ Other (please specify):

**Supporting software requirements**

Only R packages are required and they are listed in the `Libraries and dependencies used by the code` section. The citations of these packages can be found by `citation('name_of_package')` or the GitHub folder if it is not on CRAN.

**Version of primary software used** R version 4.4.1

**Libraries and dependencies used by the code** All the packages and their version numbers can be found in the following chunk of R code. Note that the instruction for settling the R environment with these packages can be found in the README in the GitHub provided below. In particular, the `renv` package will be used for this purpose. The gene-triplet analysis additionally requires Bioconductor packages `TCGAbiolinks`, `SummarizedExperiment`, and `edgeR` (installed via `BiocManager`).

```
pack_versions <- c(
  "EnvStats" = "3.0.0",
  "FNN" = "1.1.4.1",
  "IndepTest" = "0.2.0",
  "JADE" = "2.0-4",
```

```
"LaplacesDemon" = "16.1.6",
"MASS" = "7.3-65",
"Matrix" = "1.7-0",
"ProDenICA" = "1.1",
"R6" = "2.5.1",
"RColorBrewer" = "1.1-3",
"Rcpp" = "1.0.14",
"RcppArmadillo" = "14.0.2-1",
"Rdpack" = "2.6.1",
"SpatialNP" = "1.1-5",
"askpass" = "1.2.1",
"base64enc" = "0.1-3",
"boot" = "1.3-31",
"bslib" = "0.8.0",
"cachem" = "1.1.0",
"callr" = "3.7.6",
"cli" = "3.6.3",
"clue" = "0.3-65",
"cluster" = "2.1.6",
"codetools" = "0.2-20",
"colorspace" = "2.1-1",
"combinat" = "0.0-8",
"cowplot" = "1.1.3",
"cpp11" = "0.5.0",
"crosstalk" = "1.2.1",
"curl" = "5.2.3",
"dHSIC" = "2.1",
"data.table" = "1.16.2",
"desc" = "1.4.3",
"digest" = "0.6.37",
"dplyr" = "1.1.4",
"energy" = "1.7-12",
"evaluate" = "1.0.1",
"fansi" = "1.0.6",
"farver" = "2.1.2",
"fastmap" = "1.2.0",
"fontawesome" = "0.5.2",
"foreach" = "1.5.2",
"fs" = "1.6.4",
"gam" = "1.22-5",
"generics" = "0.1.3",
"ggplot2" = "3.5.1",
"glue" = "1.8.0",
"gmm" = "1.8",
"gridExtra" = "2.3",
"gsl" = "2.1-8",
"gtable" = "0.3.6",
"highr" = "0.11",
"htmltools" = "0.5.8.1",
"htmlwidgets" = "1.6.4",
"httr" = "1.4.7",
"igraph" = "2.1.2",
"isoband" = "0.2.7",
"iterators" = "1.0.14",
```

```
"jdcov" = "1.0.0",
"jquerylib" = "0.1.4",
"jsonlite" = "1.8.9",
"katlabutils" = "0.0.0.9000",
"kernlab" = "0.9-33",
"knitr" = "1.48",
"labeling" = "0.4.3",
"later" = "1.3.2",
"lattice" = "0.22-6",
"lazyeval" = "0.2.2",
"lifecycle" = "1.0.4",
"magrittr" = "2.0.3",
"memoise" = "2.0.1",
"mgcv" = "1.9-1",
"microbenchmark" = "1.5.0",
"mime" = "0.12",
"mixtools" = "2.0.0",
"multivariance" = "2.4.1",
"munsell" = "0.5.1",
"mvtnorm" = "1.3-1",
"nlme" = "3.1-166",
"nortest" = "1.0-4",
"openssl" = "2.2.2",
"pillar" = "1.9.0",
"pkgbuild" = "1.4.4",
"pkgconfig" = "2.0.3",
"plotly" = "4.10.4",
"pracma" = "2.4.4",
"processx" = "3.8.4",
"promises" = "1.3.0",
"ps" = "1.8.0",
"purrr" = "1.0.2",
"randtoolbox" = "2.0.4",
"rappdirs" = "0.3.3",
"rbibutils" = "2.3",
"renv" = "1.0.10",
"rlang" = "1.1.4",
"rmarkdown" = "2.28",
"rngWELL" = "0.10-9",
"sandwich" = "3.1-1",
"sass" = "0.4.9",
"scales" = "1.3.0",
"segmented" = "2.1-2",
"steadyICA" = "1.0",
"stringi" = "1.8.4",
"stringr" = "1.5.1",
"survival" = "3.7-0",
"sys" = "3.4.3",
"tibble" = "3.2.1",
"tidyr" = "1.3.1",
"tidyselect" = "1.2.1",
"tinytex" = "0.53",
"utf8" = "1.2.4",
"vctrs" = "0.6.5",
```

```
  "viridisLite" = "0.4.2",
  "withr" = "3.0.2",
  "xfun" = "0.48",
  "yaml" = "2.3.10",
  "zoo" = "1.8-12"
)
```

#### Parallelization used

- ☒ No parallel code used
- ☐ Multi-core parallelization on a single machine/node
  - Number of cores used:
- ☐ Multi-machine/multi-node parallelization
  - Number of nodes and cores used:

#### License

- ☒ MIT License (default)
- ☐ BSD
- ☐ GPL v3.0
- ☐ Creative Commons
- ☐ Other: (please specify)

# Part 3: Reproducibility workflow

## Scope

The provided workflow reproduces:

- ☐ Any numbers provided in text in the paper
- ☒ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))
- ☒ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below:

## Workflow

#### Location

The workflow is available:

- ☐ As part of the paper's supplementary material.
- ☒ In this Git repository: https://anonymous.4open.science/r/RJdCov-project-84EF/README.md
- ☐ Other (please specify):

#### Format(s)

- ☐ Single master code file
- ☐ Wrapper (shell) script(s)
- ☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach
- ☒ Text file (e.g., a readme-style file) that documents workflow
- ☐ Makefile
- ☐ Other (more detail in *Instructions* below)

**Instructions**

Please follow the README in the GitHub link provided above. The instructions for reproducing the simulation and real data analyses are self-contained there.

**Expected run-time**

Approximate time needed to reproduce the analyses on a standard desktop machine:

- ☐ < 1 minute
- ☐ 1-10 minutes
- ☐ 10-60 minutes
- ☐ 1-8 hours
- ☒ > 8 hours
- ☐ Not feasible to run on a desktop machine, as described here: