Author Contributions Checklist Form

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

There are two parts of data used in this paper. The main text uses synthetic data where X is generated from a multivariate student distribution and Y is either generated from a linear model or from a nonlinear one. The supplement also contains experiment results on real data, which are the datasets from UCI repository: blog feedback, concrete strength, kernel performance, news popularity, protein structure, and superconductivity; see Appendix G of Gupta et. al. (2022) Nested conformal prediction and quantile out-of-bag ensemble methods for details about these 6 datasets.

## Availability

Data **are** publicly available

Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available dat*a section, below.

### Publicly available data

Data are available online at: http://archive.ics.uci.edu/datasets

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:

### Non-publicly available data

Discussion of lack of publicly available data:

## Description

### File format(s)

CSV or other plain text: csv

Software-specific binary format (.Rda, Python pickle, etc.):

Standardized binary format (e.g., netCDF, HDF5, etc.):

Other (described here):

### Data dictionary

Provided by the authors in the following file(s):

Data file(s) is (are) self-describiing (e.g., netCDF files)

Available at the following URL:

<https://archive.ics.uci.edu/dataset/304/blogfeedback>

<https://archive.ics.uci.edu/dataset/265/physicochemical+properties+of+protein+tertiary+structure>

<https://archive.ics.uci.edu/dataset/165/concrete+compressive+strength>

<https://archive.ics.uci.edu/dataset/332/online+news+popularity>

<https://archive.ics.uci.edu/dataset/440/sgemm+gpu+kernel+performance>

<https://archive.ics.uci.edu/dataset/464/superconductivty+data>

### Additional information (optional)

All the functions including Conformalized Quantile Regression (CQR) and the methods developed in this work (EFCP and VFCP) are in the source file methods\_functions.R. All the files named …\_simulation.R uses functions from this source file and generates the data to plot, where the data is saved as …\_data\_to\_plot.RData. Finally, the …\_plot.R file loads the .RData and produces the plots.

# Part 2: Code

## Abstract

We provide several R files that contains the simulation code and the data. All the functions including Conformalized Quantile Regression (CQR) and the methods developed in this work (EFCP and VFCP) are in the source file methods\_functions.R. All the files named …\_simulation.R uses functions from this source file and generates the data to plot, where the data is saved as …\_data\_to\_plot.RData. Finally, the …\_plot.R file loads the .RData and produces the plots.

## 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 (described here):

### **Supporting software requirements**

Version of primary software used

R version 4.0.2

Libraries and dependencies used by the code

Library ggplot2 version 3.4.4

Library quantregForest version 1.3-7

Library MASS version 7.3-60

Library mvtnorm version 1.2-3

R library conformalInference, available at

library(devtools)

install\_github(repo="ryantibs/conformal", subdir="conformalInference")

### Supporting system/hardware requirements (optional)

### 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 (described here):

### Additional information (optional)

# 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 here:

## Workflow details

### 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

Each .R file we provide contains a single simulation setting in the paper, which includes for that of synthetic data and real data

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:

### Additional documentation (optional)

# Notes (optional)