JASA A&CS Author Contributions Checklist (ACC) Form

for

"Earthquake risk embedded in property prices:
Evidence from five Japanese cities"
(JASA-A&CS-2018-0253-R4)

May 4, 2021

1 Data

Abstract

Data used in this paper comprises information collected from various sources. Our main data source is the Real Estate Transaction-Price Information provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan. Other data sets include: (i) macroeconomic indicators provided by the Japan Cabinet Office, (ii) ward characteristics provided by the portal site of official statistics of Japan, and (iii) long term earthquake probabilities provided by the Japan Seismic Hazard Information System. We have simulated short term earthquake probabilities using an Epidemic Type Aftershock Sequence (ETAS) model which we have estimated using historical earthquakes provided by the Japan Meteorological Agency. The property location and earthquake risk information have been linked using the Japanese meshcode system and the geocoding service provided by Google Maps.

Availability

All data used in this paper has been generated from publicly available sources. The cleaned and compiled data sets can be found at `https://github.com/yy112/earthquake-risk` under the folder 'data'. (Source data files can be provided upon request.)

Description

The data can be briefly described as follows—our online Data Documentation that is available from `https://bit.ly/3qHcTQ3` provides many more details.

Our data set contains four separate `csv` files:

1. `individual_data.csv`

Data from various sources have been cleaned and compiled into the data file `individual_data.csv` (size: 214.7 MB). This file is compressed in the `zip` file: `individual_data.zip` (size: 11.4 MB). Each row in this file represents one property transaction record, with information on the characteristics

of the transaction, characteristics of the property, earthquake probabilities associated with the area that the property is located in, macroeconomic variables observed at the transaction period, and demographic characteristics of the ward that the property is located in. There are also interaction terms and derived variables.

2. `Xpsi-1.csv`

Short run earthquake risk data obtained through simulation has been stored in `Xpsi-1.csv` (size: 2KB). This file contains five time series (quarterly frequency) of the short run probabilities, each column corresponding to one of the five cities included in the scope of our analysis.

3. `city_range.csv`

The file `city_range.csv` (size: 1KB) contains the chosen spatial window (for the estimation of the ETAS models) corresponding to each city. Each row represents one city and the variables used in our analysis are "latMin", "latMax", "lngMin", "lngMax", corresponding to the minimum and maximum values of the latitude and longitude of the chosen spatial window (we have chosen to use rectangular spatial windows).

4. `JMA_records.csv`

Earthquake records collected by the Japan Meteorological Agency (JMA) have been stored in `JMA_records.csv` (size: 24.15 MB). Each row of this file contains one earthquake record, with information on the time and location of the earthquake. A detailed description of the records can be found in https://www.data.jma.go.jp/svd/eqev/data/bulletin/data/shindo/format_e.txt.

The data sets have been collected from the sources described below:

We use the real estate transaction price and property characteristics information provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT):

https://www.land.mlit.go.jp/webland_english/servlet/MainServlet

GDP figures are provided by the Cabinet Office website:

https://www.esri.cao.go.jp/index-e.html
⇒ https://www.esri.cao.go.jp/en/sna/menu.html (2018 August)

accessed on April 30, 2016.

CPI data are provided by the Statistics Bureau of Japan, Ministry of Internal Affairs and Communications website:

https://www.stat.go.jp/english/data/cpi/index.html.

We use the version released on August 26, 2016, which is 2015-based and integrate the monthly data into quarterly data using simple averages.

Ward attractiveness characteristics are provided by the Portal Site of Official Statistics of Japan:

https://www.e-stat.go.jp/.

The seismic intensity data are provided by the Japan Meteorological Agency (JMA):

https://www.data.jma.go.jp/svd/eqev/data/bulletin/shindo_e.html.

Long term earthquake risk (derived from the Probabilistic Seismic Hazard Maps) is provided by The Japan Seismic Hazard Information Station (J-SHIS):

http://www.j-shis.bosai.go.jp/map/JSHIS2/download.html?lang=en.

Property location and risk information are linked by using the Geocoding services provided by Google Maps:

https://developers.google.com/maps/documentation/geocoding/start.

The Cabinet Office Homepage, the Statistics Bureau, the Portal Site of Official Statistics of Japan, and the Japan Meteorological Agency permit free usage of content provided on their website. Their Terms of Use are compatible with the Creative Commons Attribution License 4.0. The J-SHIS permits free downloading and usage of data for purposes that do not violate any law.

2 Code

Abstract

The code for our analysis has been written in programming language `R` and consists of two major parts. (i) The first part concerns the estimation and simulation of the temporal ETAS models. The functions used in this part are collected in `etas_funcs.R`. (ii) The second part concerns the estimation of the multivariate error components model. The functions used in this part are collected in an `R` package `mvecr`. The manual of this package (`mvecr_manual.pdf`) provides detailed descriptions about the inputs and outputs of each function in this package. Finally, detailed code and instructions to replicate the tables and figures in this paper can be found in `replication_instructions.Rmd/pdf`.

Availability

The full `R` code for this paper (i.e., `etas_funcs.R`, the `R` package `mvecr`, and the replication master file `replication_instructions.Rmd/pdf`) is available from our GitHub repository https://github.com/yy112/earthquake-risk.

Description

`replication_instructions.Rmd` and `replication_instructions.pdf` contain the annotated code and instructions to replicate the tables and figures in the paper. The code consists of two parts (that can be executed independently):

A. Estimation and simulation of the ETAS model (`etas_funcs.R`)

- Estimation of the ETAS model, replication of the summary statistics and estimation results in Tables 48 and 50 of the Data Documentation. Computation time needed: 5 minutes.
- Simulation of short run earthquake probabilities, using the estimated ETAS parameters and a historical earthquake catalog provided by the Japan Meteorological Agency. Computation time needed: around 24 hours for 30,000 Monte Carlo runs for each city. The output of the simulation is contained in `Xpsi-1.csv`. Figures 1 and 2 of the paper can be replicated using the data provided in `Xpsi-1.csv` without having to perform the simulation first.

The following functions contained in `etas_funcs.R` are used:

- `EQcatalog`: this function generates an earthquake catalog in the format that can be used for the estimation of the ETAS model;

- `etas_estim`: this function estimates the parameters of the ETAS model using the `PtProcess` package;
- `etas_prob`: this function calculates the earthquake probability forecasts within a given time period through simulation using the `PtProcess` package.
- `gen_Xpsi_city`: this function generates the simulated earthquake probabilities for each city for the entire sample period (2006Q2–2015Q3) using `etas_prob`.

B. Estimation of the multivariate error components regression model (`R` package `mvecr`)

- Characteristics of the housing data set. Replication of Table 1 of the paper.
- Characteristics of the JSHIS long run probabilities. Replication of Table 2 of the paper.
- Main estimation results. Replication of Table 3 and Figure 3 of the paper. Computation time needed: 1–1.5 hours for each model (with a specified value for ψ).
- Sensitivity analysis regarding probability weighting functions. Replication of Table 4 and Figure 4 of the paper. Computation time needed: 1–1.5 hours for each model.
- Sensitivity analysis regarding other model specifications. Replication of Tables B1–B5 of the supplementary material. Computation time needed: 1–1.5 hours for each model with 2 error components, 3–4 hours for the 3 error components model, 40–60 minutes for the model where individual data is grouped by nearest station instead of by district.
- Importance ordering and decomposition of risk premia. Replication of Tables C6–C7 of the supplementary material. Computation time needed: 1 minute.

The following main functions contained in `mvecr` are used:

- `vectorize`: this function takes the individual records as input, groups them by the specified "time", "district", and "type" dimensions, takes the averages within each group, and stacks the results into vectors and matrices that can be used in the multiple error components regression;
- `ec_reg`: this function takes the "vectorized" data and performs maximum likelihood estimation of the multivariate error components model;
- `reg_psi`: this function takes a step further and allows one of the regressors to be transformed by a single-parameter function, with a given parameter ψ;
- `opt_psi`: this function is a wrapper function around `reg_psi` that implements a grid search to find the optimal ψ. Given a list of candidate parameters ψ, it calls the function `reg_psi` for each value of ψ on the list and records the loglikelihood value. The parameter corresponding to the highest value of the log likelihood is chosen as $\hat{\psi}$.

3 Instructions for Use

Reproducibility scope

All tables and figures in the paper can be reproduced using the four `csv` data sets and the annotated `R` code and instructions contained in `replication_instructions.Rmd/pdf` (see also Section 2 above), all of which is available from https://github.com/yy112/earthquake-risk.

Installation instructions and supporting software requirements

`R` version later than 3.6.0 is needed. Additionally the packages `dplyr` (version 1.0.0), `PtProcess` (version 3.3-13), `readr`, `R.utils`, `zoo` and `mvecr` are needed. To install the package `mvecr`, download the source file (`mvecr_0.3.0.tar.gz`) from https://github.com/yy112/earthquake-risk and choose "Install packages from Package Archive File (.zip; .tar.gz)" in `R`. More detailed installation instructions can be found in `replication_instructions.Rmd/pdf`.