`yuimaGUI`: A graphical user interface for the `yuima` package. *User Guide yuimaGUI v1.0*

Emanuele Guidotti*, Stefano M. Iacus† and Lorenzo Mercuri‡

February 21, 2017

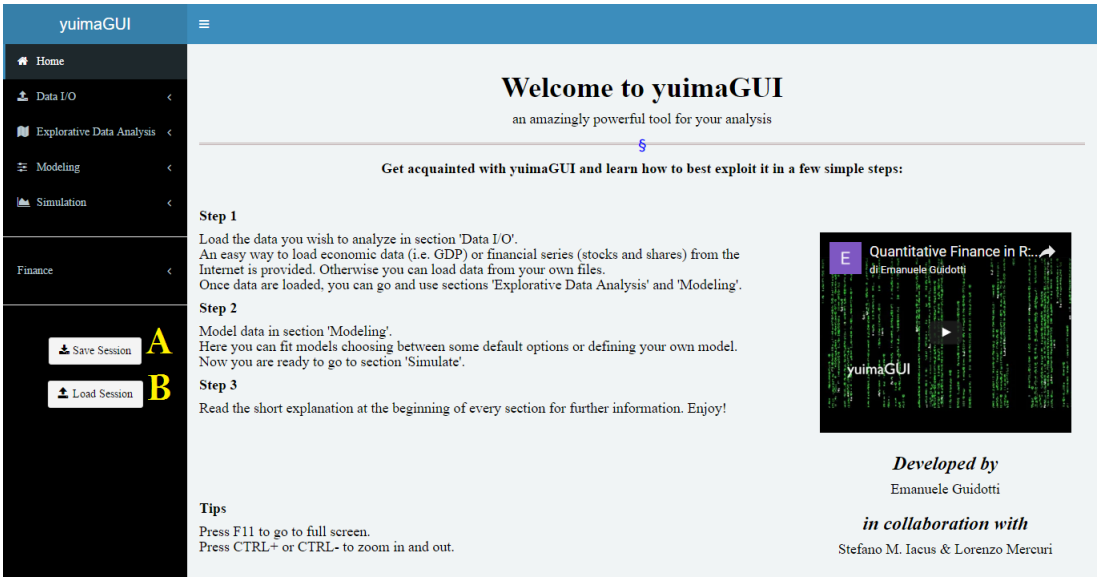
Contents

*University of Milan, `E.mail`: emanuele.guidotti@studenti.unimi.it

†Department of Economics, Management and Quantitative Methods, University of Milan, CREST Japan Science and Technology Agency, `E.mail`: stefano.iacus@unimi.it

‡Department of Economics, Management and Quantitative Methods, University of Milan, CREST Japan Science and Technology Agency `E.mail`: lorenzo.mercuri@unimi.it

1 yuimaGUI: Home



In this section it is explained how to use the interface in very few simple steps:

Step 1 Load the data you wish to analyze in section 'Data I/O'. An easy way to load economic data (i.e. GDP) or financial series (stocks and shares) from the Internet is provided. Otherwise you can load data from your own files. Once data are loaded, you can go and use sections 'Explorative Data Analysis' and 'Modeling'.

Step 2 Model data in section 'Modeling'. Here you can fit models choosing between some default options or defining your own model. Now you are ready to go to section 'Simulate'.

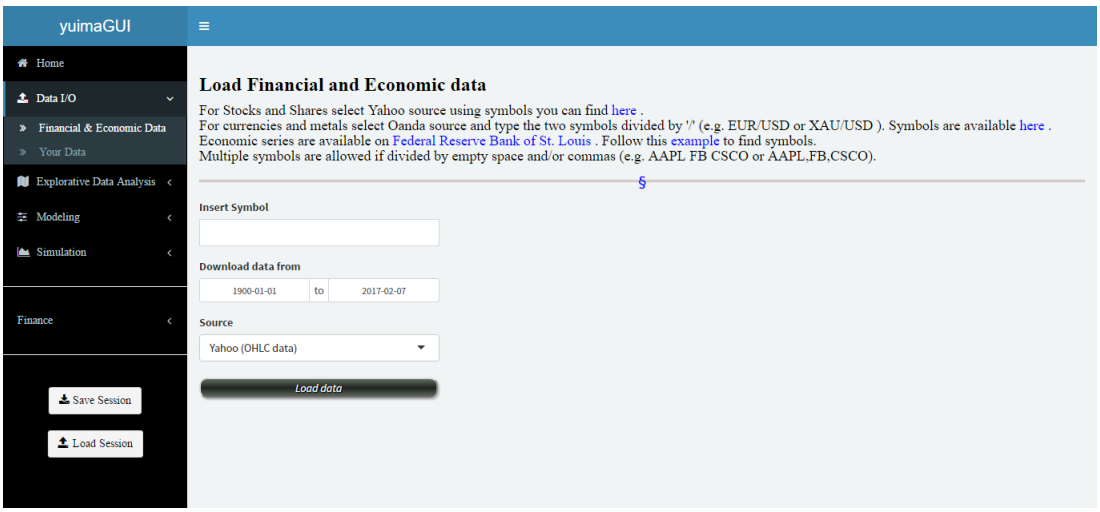
Step 3 Read the short explanation at the beginning of every section for further information. Enjoy!

On the right you can find a short video showing the usage of yuimaGUI. On the left it is the sidebar to change section. Buttons A and B allow to save everything you did using yuimaGUI (button A: *Save Session*) and to resume or share your work (button B: *Load Session*) loading a file previously saved by clicking on button A.

2 yuimaGUI: Data I/O

In this section you can load data from the Internet (section *Financial and Economic Data*) or from your local files (section *Your Data*).

2.1 Financial & Economic Data



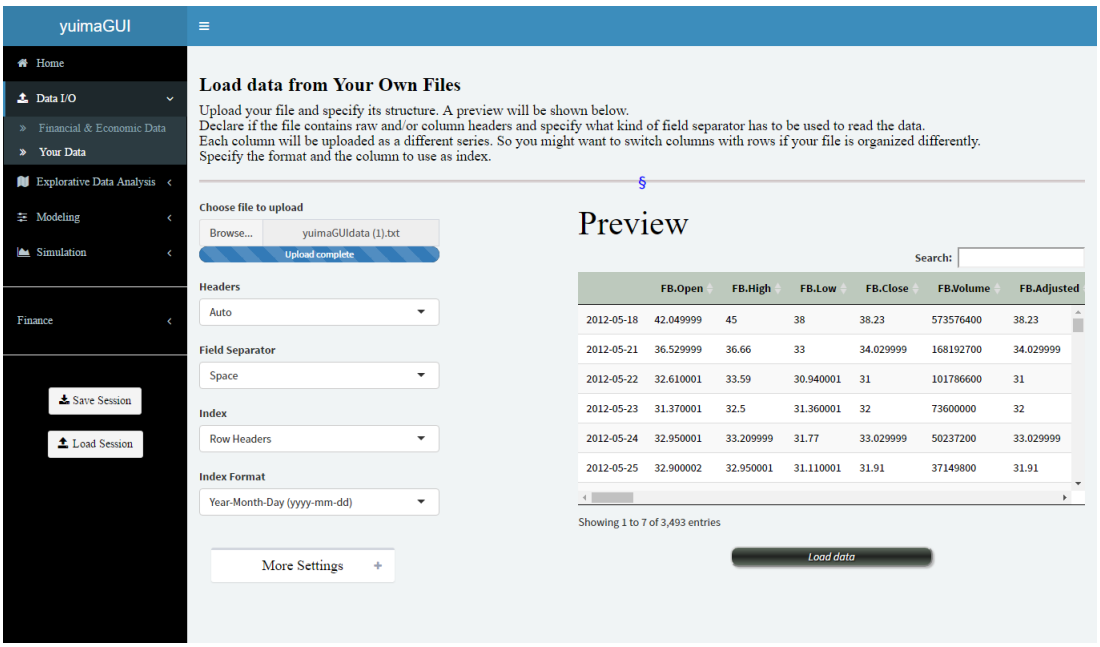
In this section you can load financial and economic data from the Internet. Type the symbols of the series you want to load, select the source of the data, choose the time horizon and click on *Load data*.
Loaded data will be displayed in the table of this section. Select a row of this table to plot the corresponding series. You can filter, save or delete the series. This table contains all the data you loaded, included those loaded in section *Your data* (see section 2.2).

Insert Symbol Type here the symbol(s) of the asset(s) you want to load. Stocks and shares, currencies and metals, economic indicators are available. All the useful links and information to find the symbols are given in the description of this section (see figure).

Download data from Select the time horizon in which you want to load the data. If data are available only after the starting date you chose, they will be loaded from when they are available.

Source Select the source for the data. If you are loading stocks and shares select *Yahoo (OHLC data)*, in this case it will be loaded a series for the opening price, the closing price, high, low, adjusted closing price and volume. If you are loading currencies and metals select *Oanda (Currencies & Metals)*. If you are loading economic indicators select *Federal Reserve Bank of St. Louis*.

2.2 Your Data



In this section you can load data from your local files. Select the file you want to upload and specify its structure. A preview will be shown on the right. Each column will be loaded as a different series. Click on *Load data* to upload the file.
Once data are loaded, it will be shown the same table you find in section *Financial and Economic Data* (see section 2.1). You can filter, save or delete the series contained in this table.

Choose file to upload Choose the file to upload.

Headers Does the file contain headers? *Auto* option tries to detect headers automatically but you should specify explicitly if the file contains headers only for rows, only for columns, both or none of them.

Field Separator Select the field separator for your file. Available options are: space, comma, semicolon, tab.

Index The index (usually time) for the series. You can use row headers as index, numeric index (1,2,3,...) or any column of your file.

Index Format If you are using a numeric index select option *numeric*, otherwise you have to specify the date format you are using.

More Settings Here you can choose the decimal and thousands separator, the string that has to be interpreted as a missing value and the line from which to start reading the file. You can eventually switch rows with columns in reading the file.

3 yuimaGUI: Explorative Data Analysis

After loading the data, you can use this section and perform various kind of analysis: change poiint estimation, clustering analysis, lead-lag and correlation analysis.

3.1 Change Point Estimation

Here you can estimate change points using different methods, both parametric and nonparametric. You can detect change points in the distribution of the increments of the process, in the distribution of the percentage increments of the process (returns for stocks and shares) and change points in the volatility of the process.

3.1.1 Nonparametric

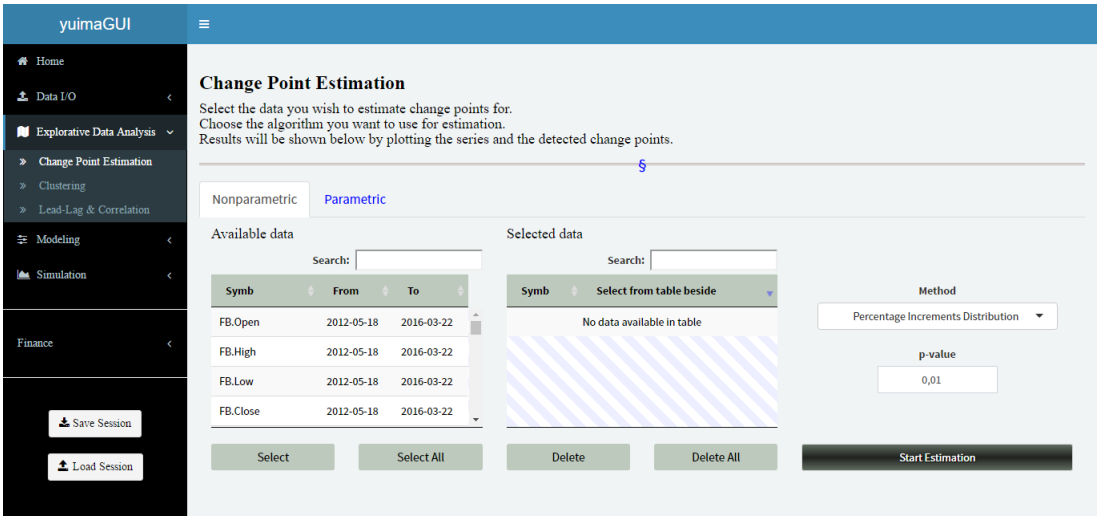


Table *Available data* contains all the series you loaded previously in section *Data I/O*. Select from this table the series you want to estimate change points for; they will be moved to table *Selected data*. Choose the method and the p-value to use for estimation and click on button *Start Estimation*. Change points will be estimated for every series in table *Selected data*.
Results will be shown below and include: plots of the series and its (percentage) increments with the detected change points; instants of change points with their p-value.

Method and p-value Available methods are *Percentage Increments Distribution* and *Increments Distribution*. They detect change points in the (percentage) increments distribution of the process, up to a confidence level given by the p-value used. The algorithm is as follows:

1. Compute the (percentage) increments of the process
2. For any t in the observation grid of the process, run a two-sample Kolmogorov-Smirnov test comparing the empirical (percentage) increments distribution between:
   - t_0 (initial time) and t
   - t and T (final time)
3. Find the minimum p-value, and the corresponding t_{min}, obtained by the KS tests. If it is greater than the **p-value** used, then the algorithm stops. Otherwise, a change point is detected at time $t_{CP} = t_{min}$.
4. Repeat point 2 and 3 but, instead of using the whole grid of observation, use only observations between t_0 and t_{CP}. Repeat again point 2 and 3 using observations between t_{CP} and T.

3.1.2 Parametric

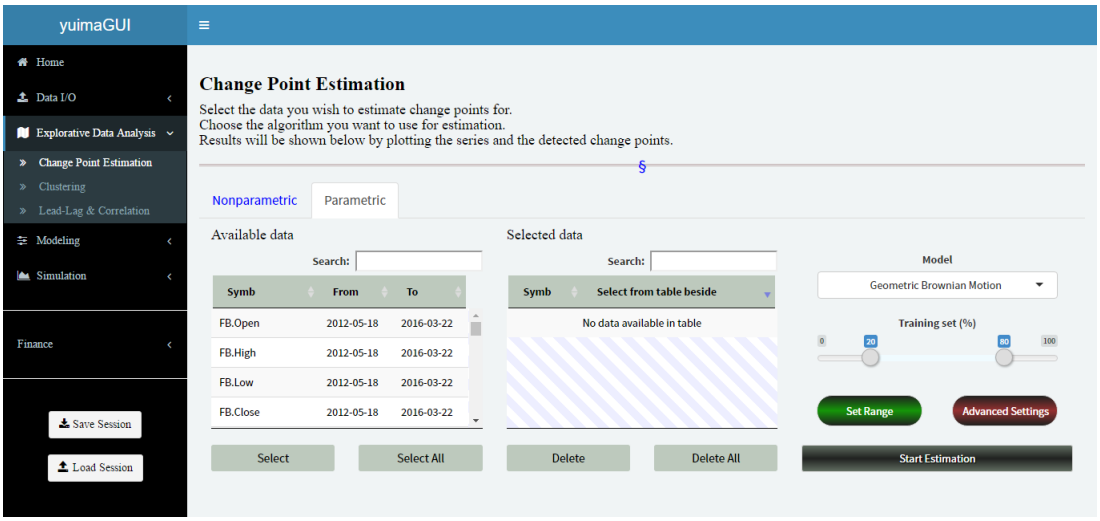


Table *Available data* contains all the series you loaded previously in section *Data I/O*. Select from this table the series you want to estimate change points for; they will be moved to table *Selected data*. Choose the diffusive model (also user-defined diffusive models can be selected, see section 4.1.2) and the training set. You can choose the window to use for change point estimation clicking on button *Set Range*. More settings are provided through button *Advanced Settings* (see section 4.1.1 for further information). Click on button *Start Estimation* and change points will be estimated for every series in table *Selected data*.
Results will be shown below and include: plot of the series with the detected change point; instant of the change point; estimates for the parameters of the chosen diffusive model before and after the change point.

Change Points are estimated using the following algorithm:

1. Estimate parameters (see section 4.1.1) for the selected **model** using the left-hand side of the **training set** (by default the first 20% of observations are used)
2. Estimate parameters (see section 4.1.1) for the selected **model** using the right-hand side of the **training set** (by default the last 20% of observations are used)
3. Estimate the change point by quasi-maximum likelihood approach using estimates obtained in point 1 and 2
4. Estimate parameters for the selected model before and after the detected change point
5. Estimate again the change point by quasi-maximum likelihood approach using estimates obtained in point 4
6. Iterate point 4 and 5 until the algorithm converges. If it does not, then an error is returned.

3.2 Clustering

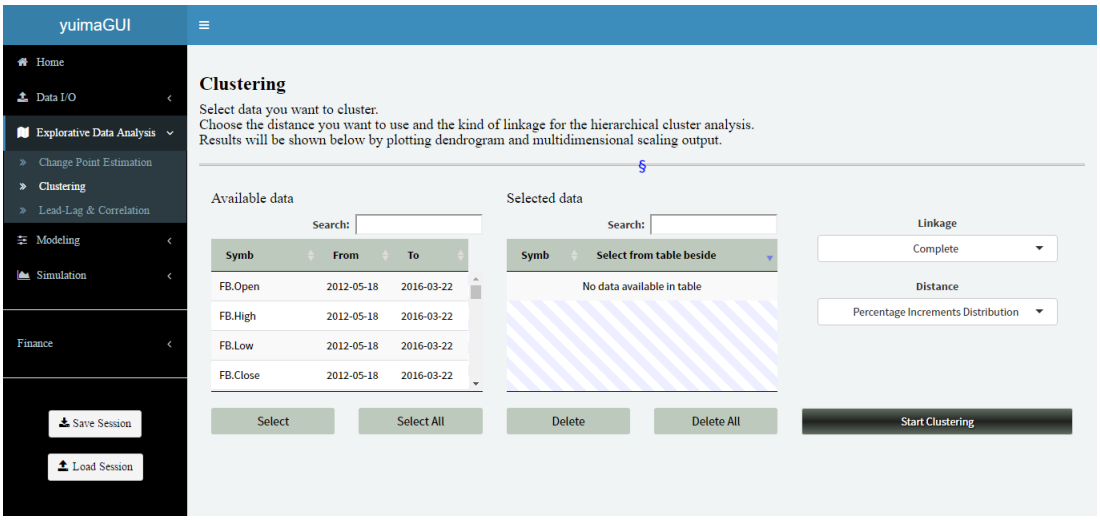


Table *Available data* contains all the series you loaded previously in section *Data I/O*. Select from this table the series you want to cluster together; they will be moved to table *Selected data*. Choose the kind of linkage and distance you want to use. Then click on button *Start Clustering*.
Results will be shown below and include both dendrogram and multidimensional scaling in order to help to identify clusters.

Linkage The kind of linkage to use for hierarchical clustering.

Distance The kind of distance to use for hierarchical clustering. Many well-known distances are provided, together with two new distances: *Percentage Increments Distribution* and *Increments Distribution*. In this case, the distance between two series is computed by:

$$d(f, g) \equiv \frac{1}{2} \int |f - g|$$

where f and g are the empirical (percentage) increments distribution of the two series.
The *Percentage Increments Distribution* is very suitable for stocks and shares because it compares the distribution of returns, thus grouping together assets with similar volatility and drift.

3.3 Lead-Lag & Correlation

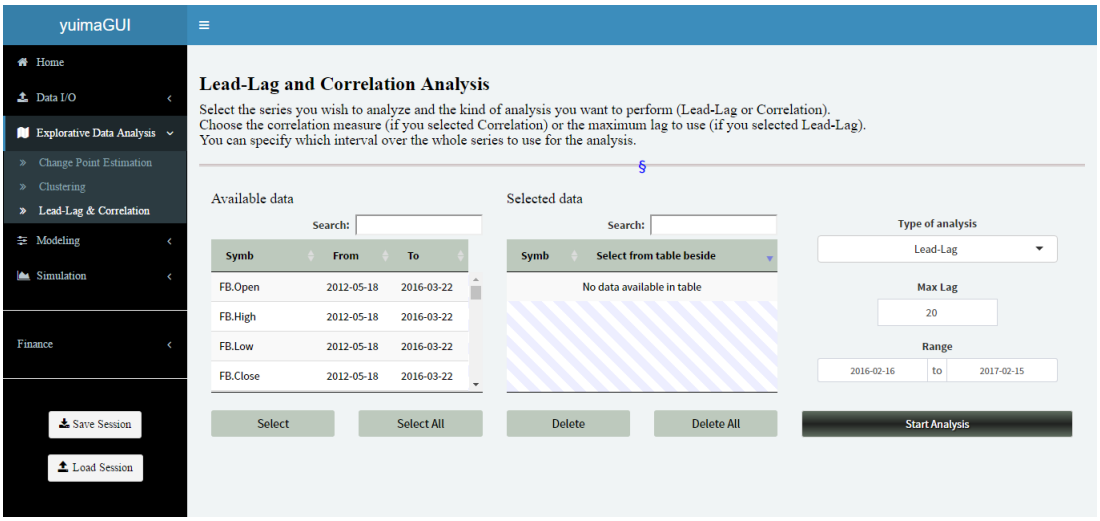


Table *Available data* contains all the series you loaded previously in section *Data I/O*. Select from this table the series you want to compute correlation or lead-lag for; they will be moved to table *Selected data*. Choose the kind of analysis to run (correlation or lead-lag) and click on button *Start Analysis*.
Results will be shown below in matrix form. If you selected *correlation* analysis then the correlation matrix will be shown. If you selected *lead-lag* analysis then the lead-lag matrix will be shown, with all the useful information on how to read the results. In this case only lead-lag effects statistically significant for the confidence level you choose are displayed.

Type of Analysis *Correlation* or *Lead-Lag*.

Max Lag This input is shown if you selected *Lead-Lag* analysis. It is the maximum lag that can be detected by the algorithm. It is expressed in days if you are using series indexed by date. It is expressed in the same unit of measure of the index if you are using numerical indexes.

Method This input is shown if you selected *Correlation* analysis. It is the method to use to compute the correlation matrix.

Range Select the window of observation to use to carry out the analysis.

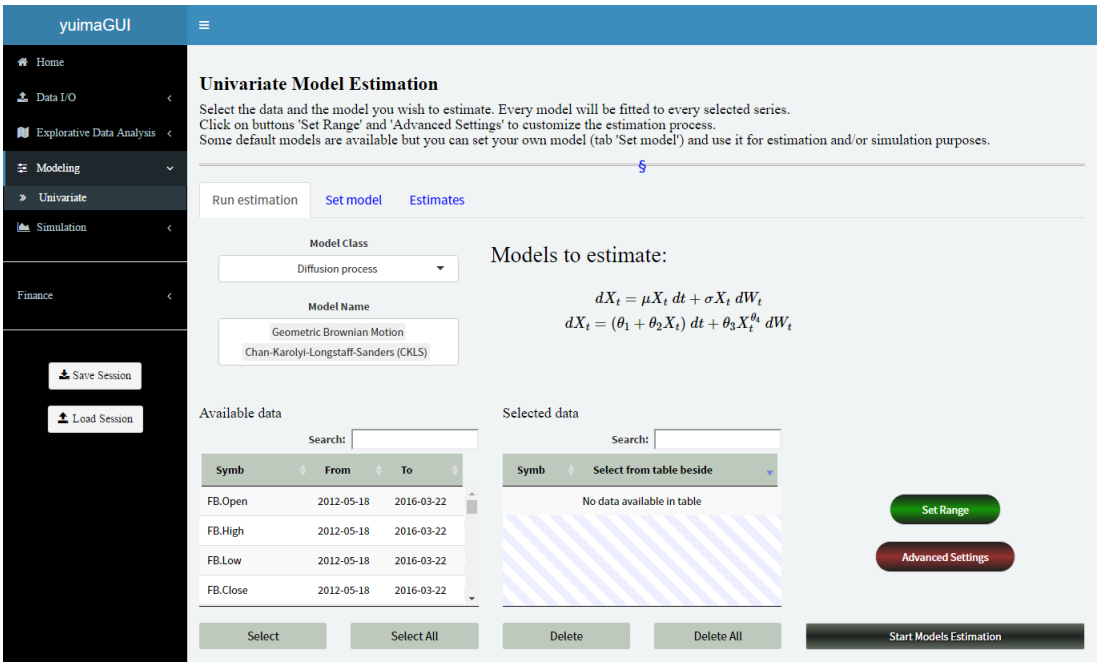
4 yuimaGUI: Modeling

At the moment, only univariate models are available.

4.1 Univariate

In this section you can estimate univariate models on data previously uploaded (*Run estimation*, see section 4.1.1) and define your own model to estimate or simulate (*Set Model*, see section 4.1.2). Finally, diagnostic tools for goodness of fit are provided (*Estimates*, see section 4.1.3).

4.1.1 Run estimation



Select the class of the model you want to estimate. Then, select the model(s) to estimate. For every class, you can choose the model among a list of default models and the ones you set in section *Set Model* (see section 4.1.2). Table *Available data* contains all the series you loaded previously in section *Data I/O*. Select from this table the series you want to estimate the models on; they will be moved to table *Selected data*. You can choose the window of observation to use for estimation clicking on button *Set Range*. More settings are provided through button *Advanced Settings*. Click

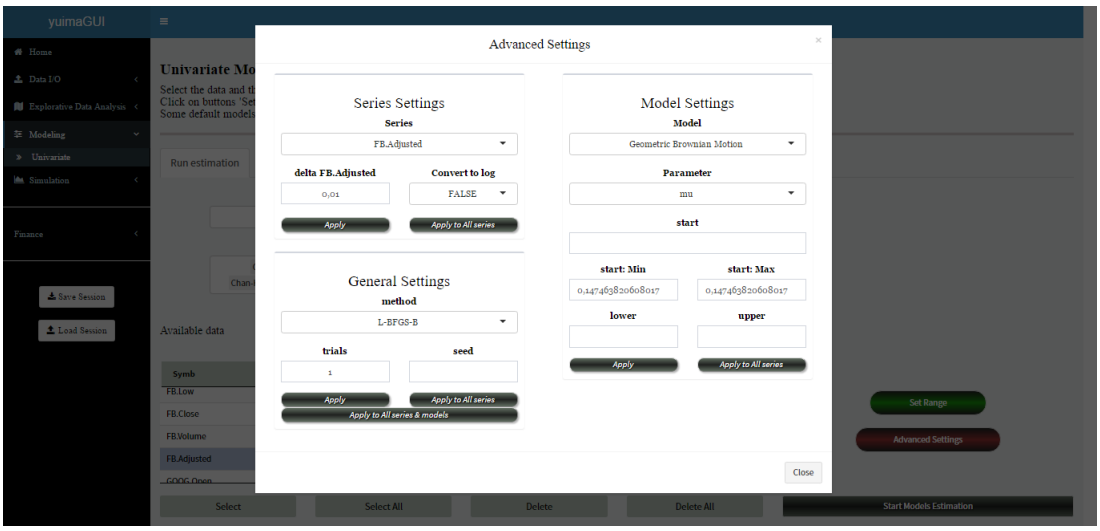
on *Start Models Estimation* and every selected model will be fitted on every selected series. Results will be shown in section *Estimates* (see section 4.1.3).

Model Class The class of the model to estimate. At the moment, available options are: *Diffusion Process*, *Fractional Process*, *Compound Poisson*, *CARMA* (Continuous Time ARMA model), *COGARCH* (Continuous Time GARCH model).

Model Name The model to estimate. You can choose among default models and those set in section *Set Model* (see section 4.1.2).

Set Range Here you can set, for every series you selected, the observation window to use for model estimation.

Advanced Settings Here you have full control over the estimation settings. Estimation is carried out via quasi-maximum likelihood approach. First, a random starting value is generated for every parameter and we try to maximize the likelihood starting from these random values. We do this for a number of times given by the input **trials**. Finally, we select the maximum likelihood estimates choosing among the optimization problems we ran. Of course, the higher the number of **trials** the more reliable the estimates (but also more time consuming the algorithm).



Series Settings: Series. Select a series for which to modify the estimation settings.

Series Settings: delta. Distance between observations. If there are 252 observation per year (as in financial series) and you want to measure time in years, then you should choose $delta = 1/252 \simeq 0.004$. If the selected series has numerical indexes then it is recommended to use the true distance between observations (done by default) in order to have the estimates for the parameters using the same units of measure of the series. For many default models it will be possible to convert the estimates in various unit of measure (see section 4.1.3).

Series Settings: Convert to Log. Convert series to logarithm? This option is available only for positive series. If you select *True* then the model(s) will be estimated on the logarithm of the series. If you simulate a model estimated this way (see section 5.1.1) then the logarithm of the trajectory will be simulated and finally it will be converted back to the original trajectory computing its exponential.

Series Settings: Apply. Apply the *Series Settings* changes to the selected series.

Series Settings: Apply to All Series. Apply the *Series Settings* changes to all series.

Model Settings: Model. Select a model for which to modify the estimation settings

Model Settings: Parameter. Select a parameter of the selected model

Model Settings: start. Starting value of the selected parameter in the optimization problem. If not specified, it is generated randomly between **start Min** and **Start Max**.

Model Settings: start Min. The lower bound to randomly generate the starting value for the selected parameter in the optimization problem.

Model Settings: start Max. The upper bound to randomly generate the starting value for the selected parameter in the optimization problem.

Model Settings: lower. Lower bound for the estimate of the selected parameter.

Model Settings: upper. Upper bound for the estimate of the selected parameter.

Model Settings: Apply. Apply the *Model Settings* changes in the estimation of the selected model only on the selected series.

Model Settings: Apply to All Series. Apply the *Model Settings* changes in the estimation of the selected model on all the selected series.

General Settings: method. The method to use in maximizing the likelihood.

General Settings: trials. The number of times to try optimizing the likelihood.

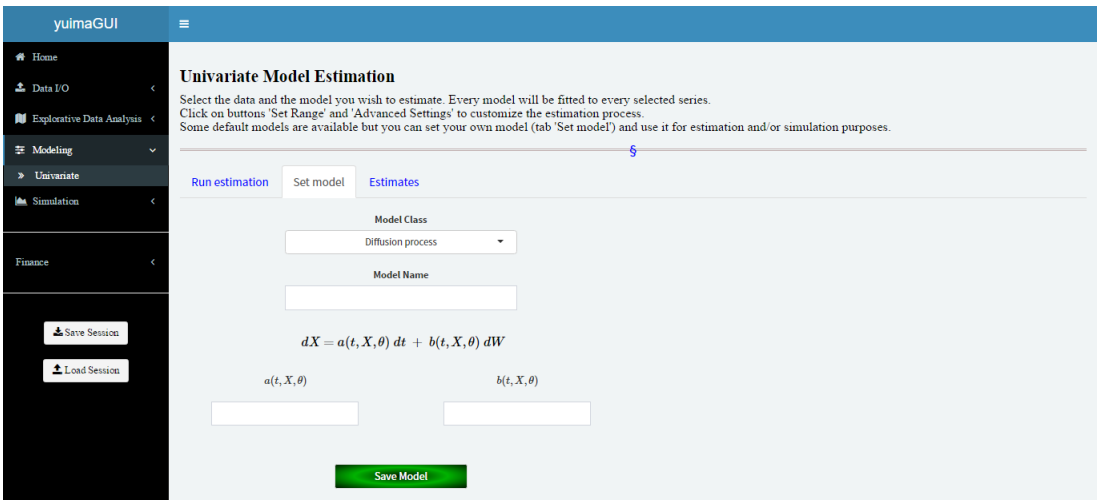
General Settings: seed. The random number generator seed. It can be unspecified.

General Settings: Apply. Apply the *General Settings* changes in the estimation of the selected model only on the selected series.

General Settings: Apply to All Series. Apply the *General Settings* changes in the estimation of the selected model on all the series.

General Settings: Apply to All Series & Models. Apply the *General Settings* changes in the estimation of all the models on all the series.

4.1.2 Set Model



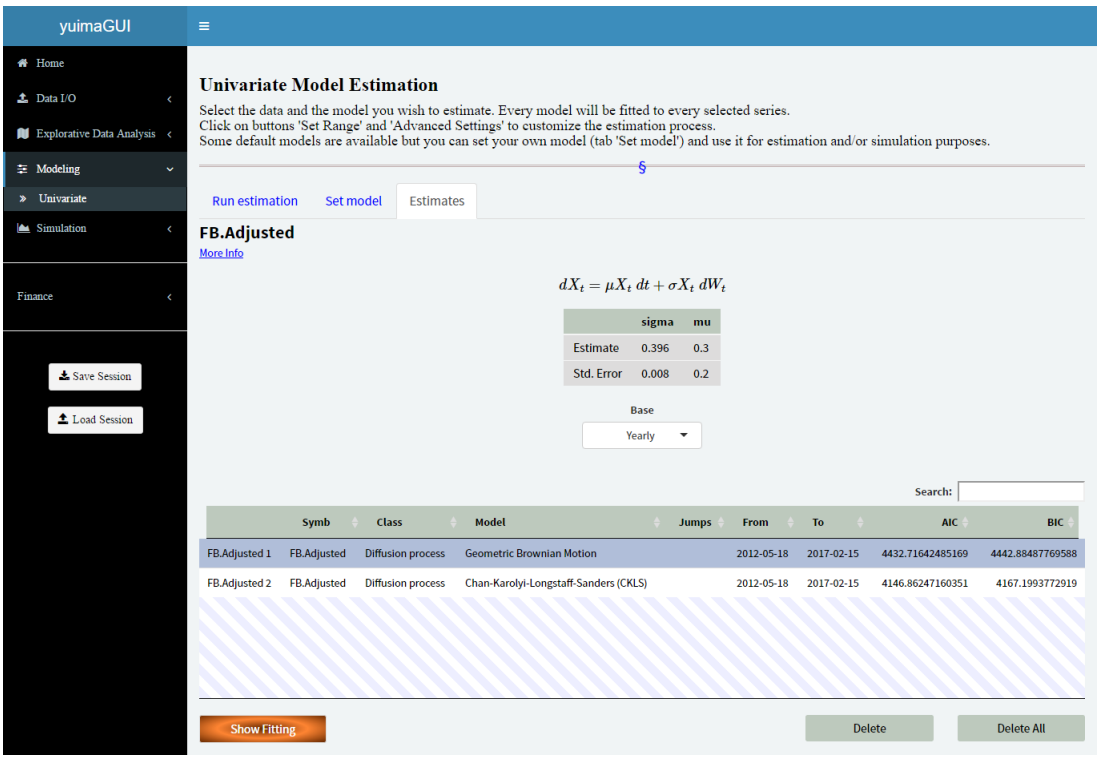
In this section you can define your own models to use for estimation and simulation purposes. Diffusion process you set here will be available also for change point analysis (see section 3.1.2). Select the class of the model to define and give a name to your model. Then define the model and click on *Save Model*.

Model Class Select the class of the model to define. Available options are: *Diffusion process*, *Fractional process* and *Compound Poisson*.

Model Name Type a name for your model.

Definition of Model For diffusion and fractional process you can specify the drift and diffusion terms. For compound poisson process you can specify the intensity. Type into the specific boxes the expression for the required terms. You can use standard functions (like *cos, sin, log, exp*, ...). Text strings are interpreted as parameters but x denotes the value of the process and t denotes the value of the index (usually time, but not necessarily).

4.1.3 Estimates



Here you can find all the information about the estimation you ran (see section 4.1.1). Estimated models are stored in the table you can see in the picture above, which includes also some information criteria (AIC and BIC). Click on a row of the table and the estimates will be shown. Click on the little link *More Info* on the upper-left corner of the tab to show the advanced settings used for estimation. The orange button on the bottom-left corner of the tab provides a very useful tool to visualize the goodness of fit of the model.

Base This input is available only if you are using series indexed by date. It controls the unit of measure of the estimates shown in the table above (yearly, semestral, ..., daily). If you are using numerical indexes then estimates are given in the same unit of measure used by the series the model was estimated on. The conversion is not always possible: user defined models (see section 4.1.2) are not supported while the majority of default models are supported. If the conversion is not possible, then estimates obtained with the chosen **delta** (see section 4.1.1 → Advanced Settings → Series Settings: delta) will be shown.

More Info It shows the advanced settings used for estimation and displays the estimates without applying any conversion, thus related to the **delta** used for estimation (see section 4.1.1 → Advanced Settings → Series Settings: delta).

Show Fitting This tool is available for: *Diffusive*, *Compound Poisson* and *COGARCH* models.

Diffusive Process. It is shown the plot of the empirical versus the theoretical distribution (standard gaussian) of standardized increments:

$$\Delta W_t = \frac{\Delta X_t - \hat{\mu}(t, X_t)\Delta t}{\hat{\sigma}(t, X_t)}$$

where the notation is the same of the following general diffusive process:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$$

It is also given the p-value for the Kolmogorov-Smirnov test that checks if the empirical and theoretical distribution are the same.

Compound Poisson. It is shown the plot of the empirical versus the estimated cumulative intensity of the process. It is also shown the plot of the empirical versus the estimated distribution of jumps (increments of the process) and the p-value for the Kolmogorov-Smirnov test that checks if the two distribution are the same.

COGARCH. It is shown the plot of the empirical versus the estimated volatility process.

5 yuimaGUI: Simulation

At the moment, only univariate simulation is available.

5.1 Univariate

In this section you can simulate univariate models, both previously estimated (*Estimated models*, see section 5.1.1) and non-estimated (*Non-estimated models*, see section 4.1.2). Trajectories, distributions and quantiles will be computed and readily available (*Simulations*, see section 4.1.3).

5.1.1 Estimated models

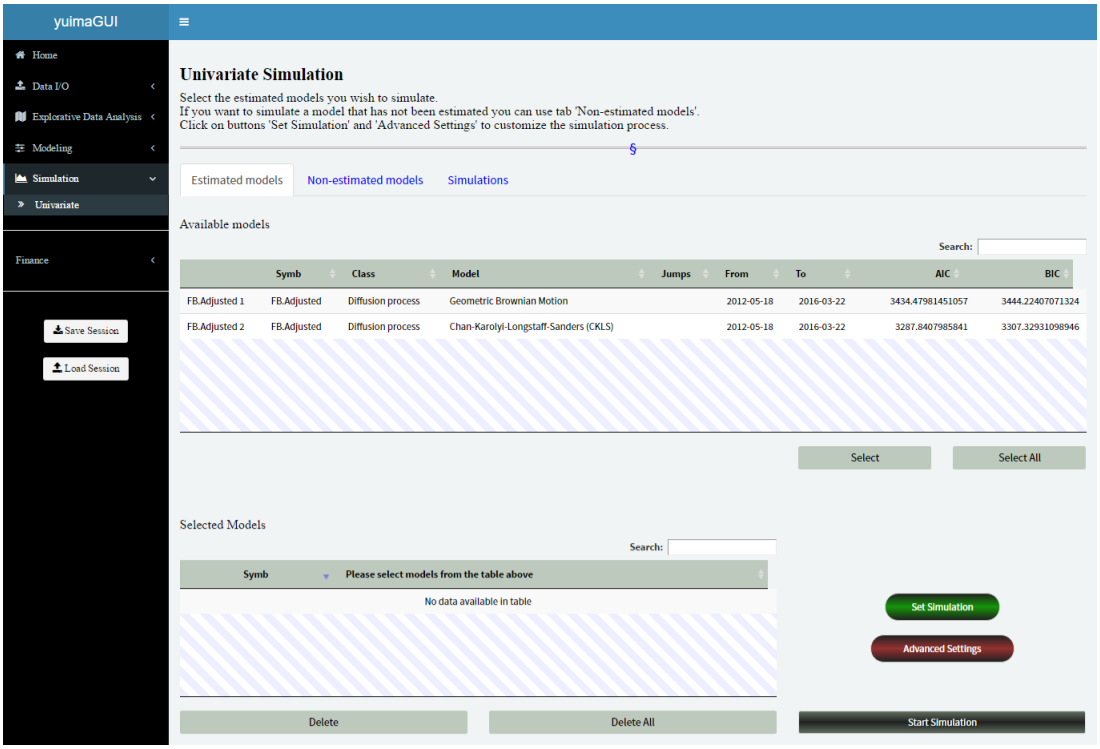


Table *Available Models* contains all the models you estimated previously in section *Modeling* → *Univariate* (see section 4.1.3). Select from this table the model(s) you want to simulate. They will be moved to table *Selected Models*. You can set the simulation clicking on button *Set Simulation*. More settings are provided through button *Advanced Settings*. Click on *Start Simulation* and every selected model will be simulated.
Results will be shown in section *Simulations* (see section 5.1.3).

Set Simulation Here you can set the:

Simulation Interval Range for the simulation.

Initial Value Initial value for the process to simulate. By default, it is the last value of the series the model was estimated on.

Number of Simulations Number of trajectories to simulate.

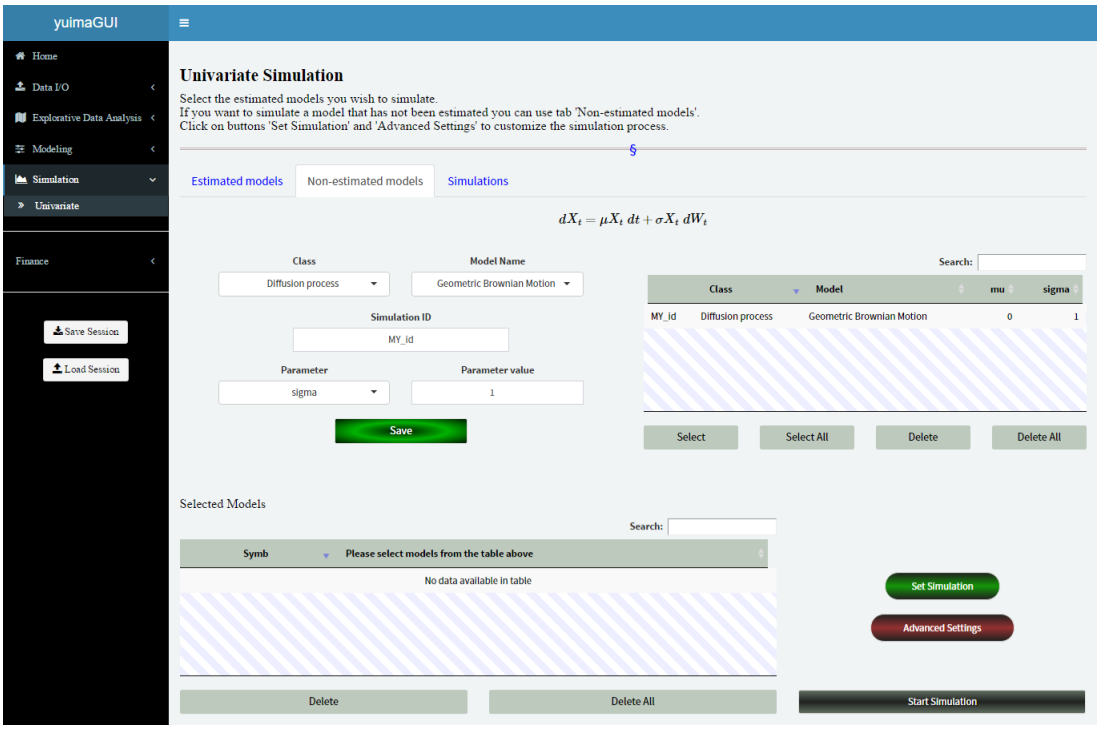
Number of Steps Number of steps per trajectory. By default, it is computed using the same observation frequency of the series the model was estimated on.

Advanced Settings Here some advanced options are provided:

RNG seed The Random Number Generator seed. By default, it is not specified.

Save trajectory Save the whole trajectory or only its final value? Default to TRUE: save the whole trajectory. It is recommended to set it to FALSE when running a high **number of simulations** and/or when using a high **number of steps** per trajectory; otherwise the computation may be extremely slow.

5.1.2 Non-estimated models



Here you can simulate model(s) specifying the values for its parameters. Select the class of the model to simulate (*Diffusion process, Fractional process, Compound Poisson*), the name of the model and give an ID to your simulation. Then select a parameter of the chosen model and set a value for it. Click on the *Save* button and the model will appear in the table beside. Select and set a value for all the parameters of the model. Then you can select the model and it will appear in table *Selected Models*. You can set the simulation clicking on button *Set Simulation*. More settings are provided through button *Advanced Settings*. Click on *Start Simulation* and every selected model will be simulated.
Results will be shown in section *Simulations* (see section 5.1.3).

Set Simulation Here you can set the:

Simulation Interval Range for the simulation.

Initial Value Initial value for the process to simulate.

Number of Simulations Number of trajectories to simulate.

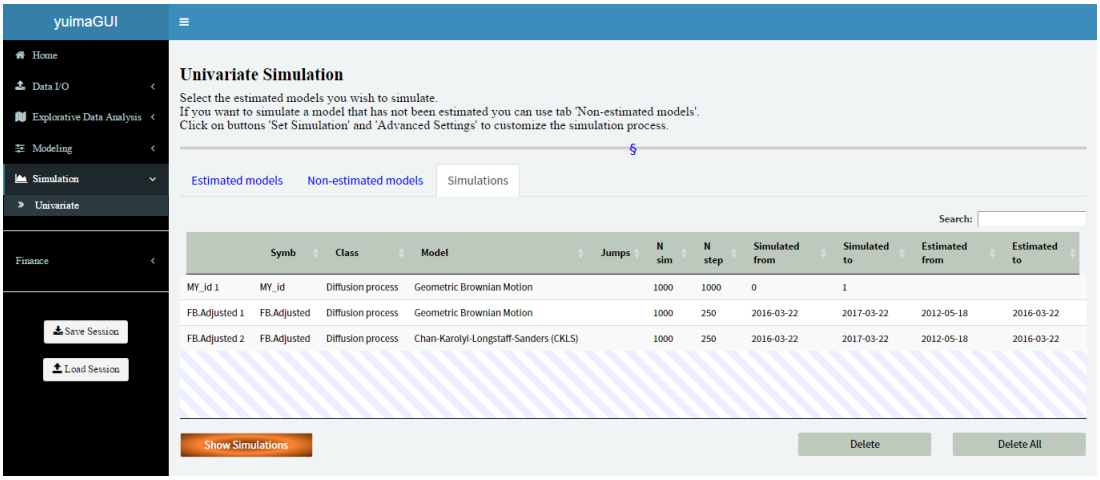
Number of Steps Number of steps per trajectory.

Advanced Settings Here some advanced options are provided:

RNG seed The Random Number Generator seed. By default, it is not specified.

Save trajectory Save the whole trajectory or only its final value? Default to TRUE: save the whole trajectory. It is recommended to set it to FALSE when running a high **number of simulations** and/or when using a high **number of steps** per trajectory; otherwise the computation may be extremely slow.

5.1.3 Simulations



Here you can find all the information about the simulations you ran (see section 5.1.1 and 5.1.2). Simulations are stored in the table you can see in the picture above. Click on the orange button on the bottom-left corner to show simulations.

Show Simulations Here trajectories, distribution and quantiles are shown in an very user-friendly way. The first plot (see picture below) shows the simulated trajectories while the second one shows the distribution of the final value of the trajectories. You can click on any point in the first plot and the second one will show the distribution of the trajectories at that point. You can adjust the number of bins of the histogram and compute quantiles with the sliders beside. You can also save and download in a text file both the trajectories and the distributions shown.

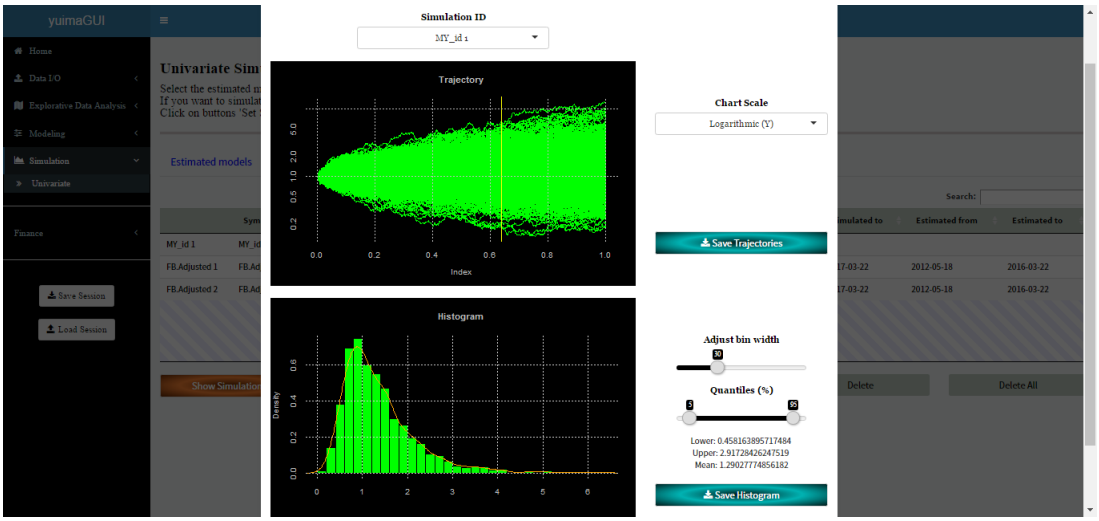


Chart Scale Scale for plotting trajectories: *Linear, Logarithmic(Y), Logarithmic(X), Logarithmic(XY)*

Save Trajectories Download a text file with all the simulated trajectories

Adjust bin width Change the number of bins used by the histogram

Quantiles (%) Select the lower and the upper percentage quantile. Below it is shown the values of the quantiles and the mean of the distribution between the two quantiles. This tool is very useful to easily compute risk measures such as Value at Risk and Expected Shortfall.

Save Histogram Download a text file with all the information about the histogram

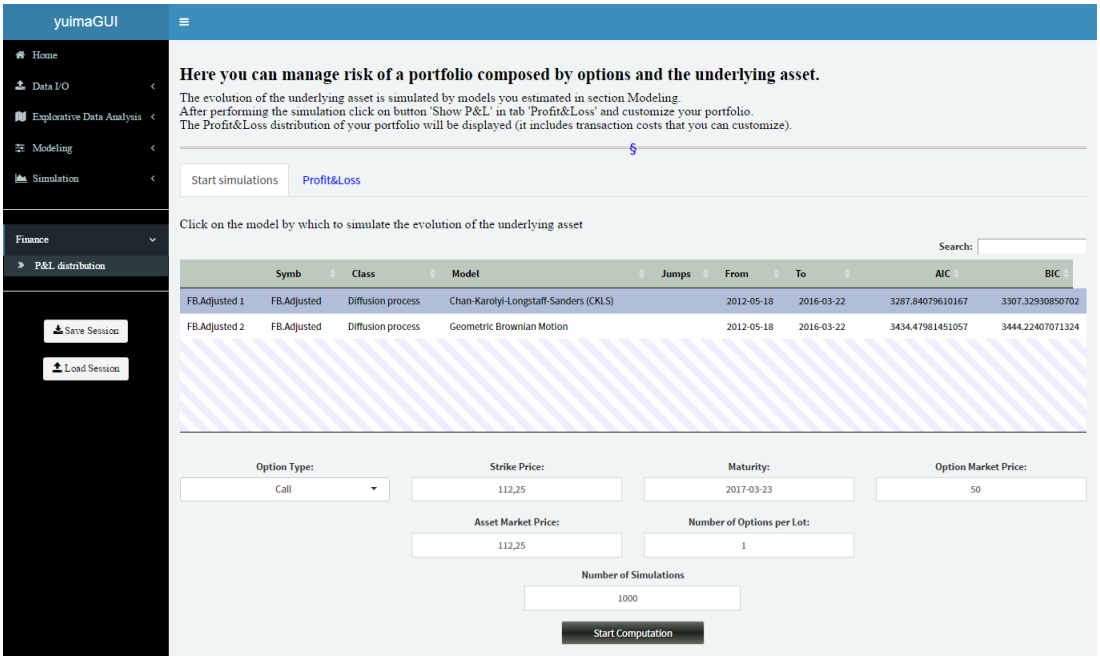
6 yuimaGUI: Finance

Here you can find useful tools specifically designated for Finance.

6.1 P&L distribution

In this section you can simulate the Profit and Loss distribution of a portfolio composed by european call/put options and the underlying asset; trading costs are included and customizable. The evolution of asset prices is simulated by models you estimated in section *Modeling* $\rightarrow$ *Univariate* (see section 4.1). After running the simulation (*Start simulations*, see section 6.1.1), results will be shown in section *Profit&Loss* (see section 6.1.2). Here you can choose the number of options and assets to include in your portfolio and the Profit&Loss distribution will be plotted. Quantiles and risk measure are provided in a very user-friendly way.

6.1.1 Start simulations



The table contains models you estimated previously in section *Modeling* → *Univariate* (see section 4.1.3). Select the model to simulate the evolution of the asset. Trajectories are simulated from the end of the series the model was estimated on to the maturity of the option. Then, specify the parameters of the option, the last price of the asset, the number of trajectories to simulate and click on *Start Computation*.
Results will be shown in section *Profit&Loss* (see section 6.1.2).

Option Type, Strike Price, Maturity Type (*Call* or *Put*), strike price and maturity of the European option

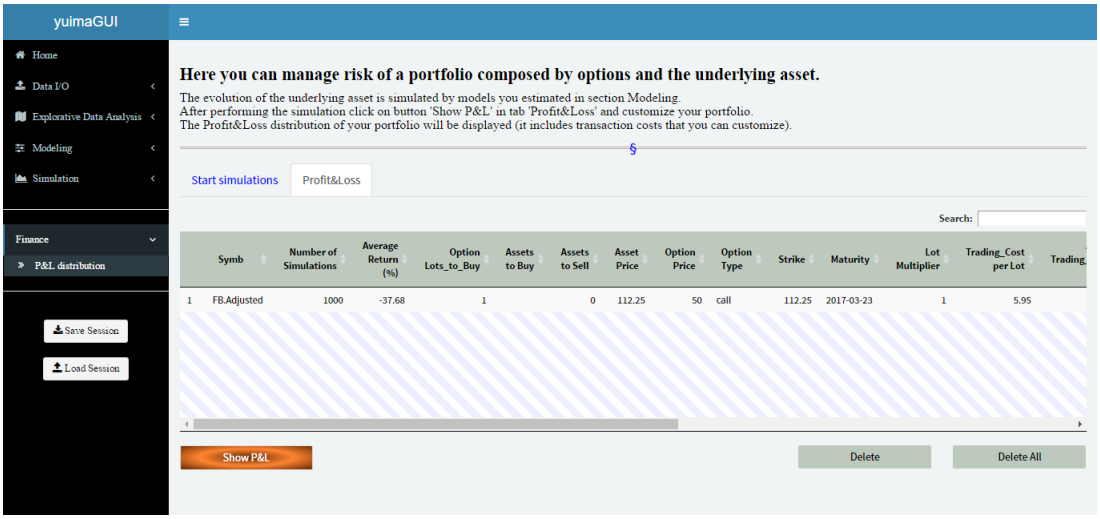
Option Market Price The market price of the option. You should type here the Ask price because you are simulating to buy the option (no short selling on options are allowed).

Asset Market Price The market price of the asset. By default it is the last price of the series the model was estimated on.

Number of Options per Lot Options are not usually traded individually but in Lots. Type here how many options a Lot includes.

Number of Simulations The number of trajectories to simulate.

6.1.2 Profit&Loss



Here you can find all the information about your simulated portfolios (see section 6.1.1). Portfolios are stored in the table you can see in the picture above. Click on the orange button on the bottom-left corner to show Profit&Loss distributions, adjust the portfolios, manage risk and customize trading costs.

Show P&L The histogram (see picture below) shows the Profit&Loss distribution of the portfolio composed by the number of options and assets you type into inputs **Option: number of Lots** and **Number of Assets to Buy/Sell**. You can adjust the bin width and compute quantiles and risk measures using the slides beside. You can also modify on the fly the parameters of the option and trading costs expanding boxes **Modify Option** and **Trading Costs**. The Profit&Loss distribution will be readily updated. You can save the portfolio clicking on button *Save Changes*.

Option: Number of Lots. Number of option Lots to buy.

Number of Assets to Buy/Sell. If you are buying a Call option, this input represents the number of assets to sell to hedge the portfolio. If you are buying a Put option, this input represents the number of assets to buy to hedge the portfolio.

Adjust bin width Change the number of bins used by the histogram

Quantiles (%) Select the lower and the upper percentage quantile. Below it is shown the values of the quantiles and the mean of the distribution between the two quantiles. This tool is very useful to easily compute risk measures such as Value at Risk and Expected Shortfall.

Save Changes Save the portfolio into the table of this section (*Profit&Loss*, section 6.1.2)

Modify Option Expand this box to modify the type, strike price and market price of the option. The Profit&Loss distribution will be readily updated.

Trading Costs Expand this box to modify the trading costs: percentage commission on buying/selling assets, minimum commission on buying/selling assets, yearly interest rate on short selling (asset), fix commission on buying one Lot of options. The Profit&Loss distribution will be readily updated.