Eco-RETINA — Quick Start Guide v1.13

1. What is Eco-RETINA?

Eco-RETINA is an innovative, environmentally friendly algorithm designed for out-of-sample prediction. It operates as a flexible, regression-based approximator — linear in parameters but nonlinear in inputs. The algorithm uses a selective model search to enhance performance while maintaining interpretability. Currently, Eco-RETINA supports cross-sectional data and a pooled panel. It also supports some time series models.

2. Software and Graphical User Interface (GUI)

Eco-RETINA is developed in Python and Jupyter, is open source, and is available on GitHub.

The Graphical User Interface (GUI) makes the software easy to use:
- No programming skills required
- Intuitive and user-friendly design
- Simple installation and operation

3. System Requirements

- Operating system: Windows 10 or 11 or Mac.
- Memory requirement: 170 MB for Windows.
- Download link for Windows: https://github.com/jcapilla780/Eco-RETINA
- Mac installer attached

4. Installing the GUI

Download the latest GUI release from:
https://github.com/jcapilla780/Eco-RETINA/releases/tag/v1.0.0

Mac installer attached.

Follow these steps:
1. Run the file EcoRETINA.exe (no installation needed). 2. If Windows shows a security warning, click: More info → Run anyway.
3. This is a standalone executable; it doesn't require a Python installation.
4. The program will automatically create a desktop icon after the first launch.

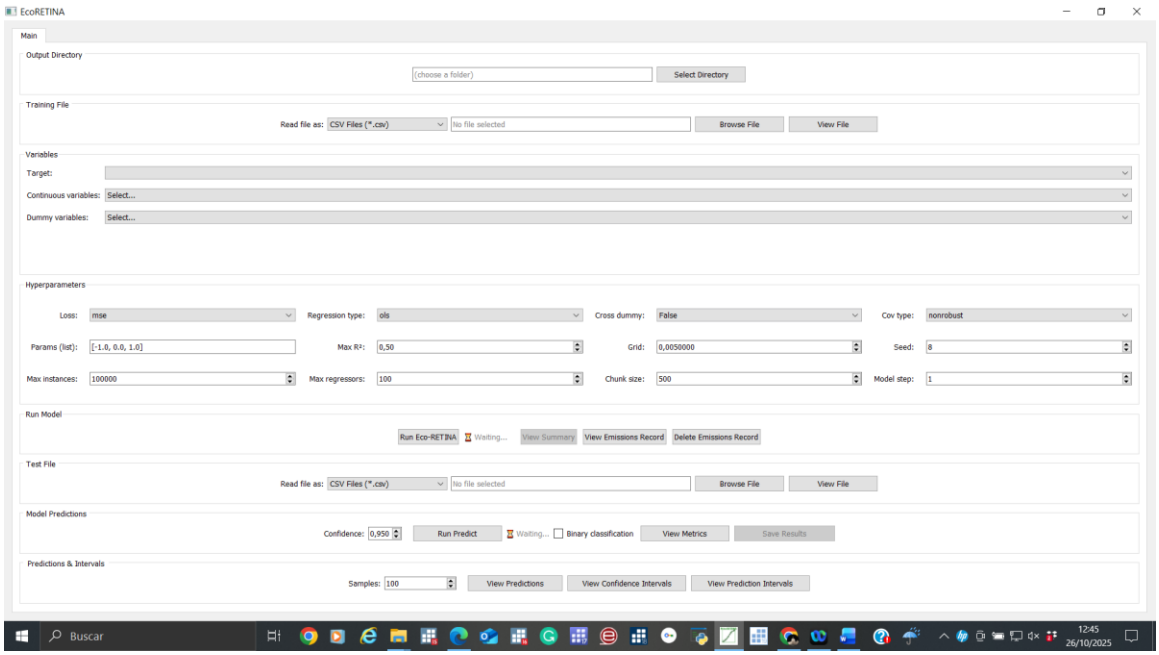
Tip: If the download is blocked, try the following steps:

A. Verify whether your antivirus software is preventing the download.
B. Bypass or override any Windows security warnings that appear.

C. Copy the release URL into your web browser and manually download the `.exe` file, confirming the download even if it is flagged as “unsafe.”

You are now ready to use Eco-RETINA—next, the first dialog box.

We provide a simple example of a data set, input, and output below.



5. Running Eco-RETINA

1. Double-click the Eco-RETINA desktop icon.
2. Select a directory, file format, and training dataset. The .csv format is convenient to use.
3. Check your input file — variable names must be listed in the first row.
4. Choose your target variable.
5. Select the variables that are continuous and numerical.
6. Select the binary (dummy) variables.
7. Keep the default settings for the remaining options for the moment.
8. Click Run Eco-RETINA and wait for completion.
9. Click View Summary to inspect the results.

6. Output and Interpretation

After execution, you can:

-Discover the relevant transformations of the inputs and formulate the forecasting model
- Examine model diagnostics and coefficient estimates
- Compare results with alternative estimation techniques
- Modify the model, perform forecasts. and graph forecasts and confidence intervals

7. Out-of-Sample Forecasting

To perform an out-of-sample conditional forecast, please provide a test file with the same variable structure as the training data, then select:

1. Run Predict
2. View Metrics
3. Predictions and Intervals

8. Example Dataset

For illustration, Eco-RETINA includes the Wooldridge Wagepan dataset, widely used in econometrics for model testing and comparison. We provide two files: one containing the first three quarters of the observations (Wagepantrain) and another with the last quarter of the sample (Wagepantest). Both files contain the same variables. The description of the variables follows.

Variables marked with *** are continuous.

Variable	Description
year	Year of observation (1980–1987)
agric	1 = Works in agriculture
black	1 = Individual is Black
bus	1 = Works in the business/finance sector
construc	1 = Employed in construction
ent	1 = Employed in entertainment services
exper***	Years of potential labor market experience
fin	1 = Employed in finance, insurance, or real estate
hisp	1 = Individual is Hispanic
poorhlth	1 = Reports poor health
hours***	Average weekly hours worked
manuf	1 = Employed in manufacturing
married	1 = Married
min	1 = Employed in mining
nrthcen	1 = Lives in North Central region

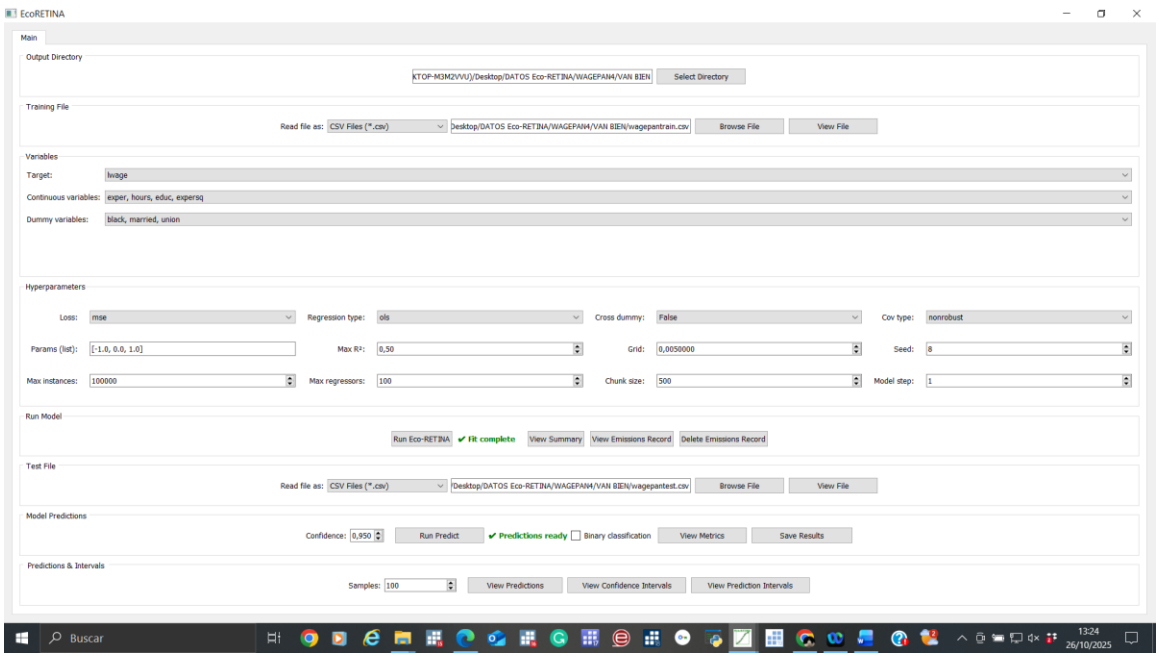
nrtheast	1 = Lives in Northeast region
occ1–occ9	Occupation dummies (categories 1–9)
per	1 = Employed in personal services
pro	1 = Employed in professional services
pub	1 = Public sector employee
rur	1 = Lives in rural area
south	1 = Lives in the South
educ***	Years of education completed
tra	1 = Employed in transportation
trad	1 = Employed in trade sector
union	1 = Union member
lwage	Logarithm of hourly wage
d81–d87	Year dummies (1981–1987; 1980 = base year)
expersq***	Square of experience (exper^2, nonlinear effect)

9. Sample Input and Output

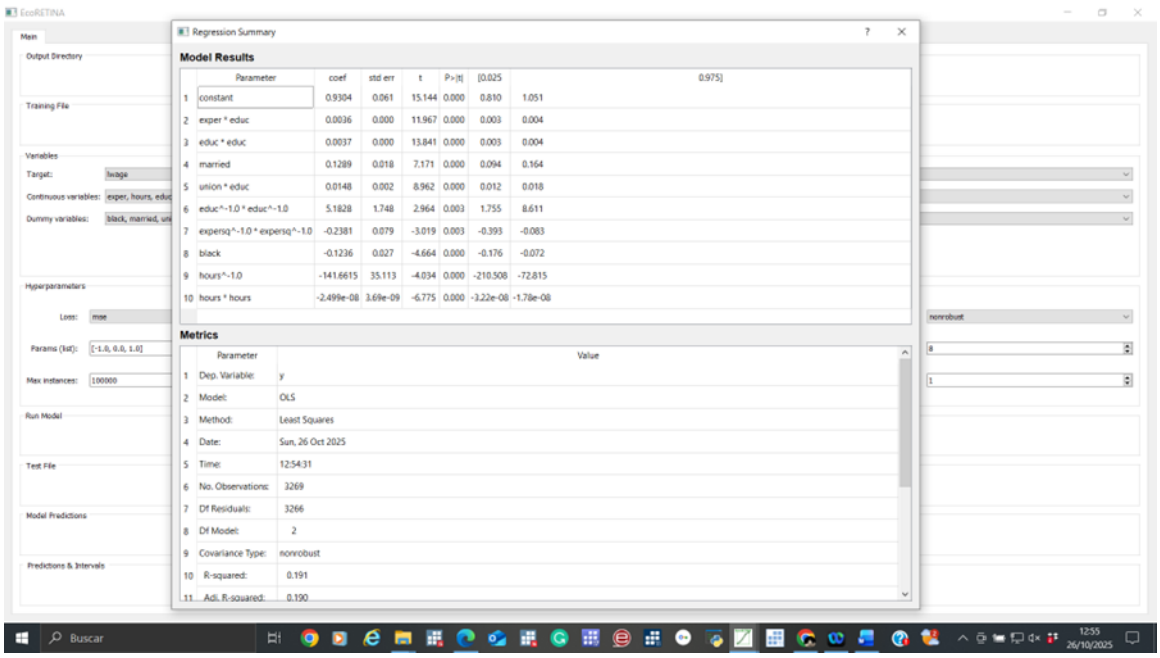
When you load the example data (Wagepantrain), you can see a preview similar to the Wagepan dataset. This dataset is about the salaries of young workers in a part of the United States. You can use lwage as the target variable.

You can use the four continuous variables, exper, educ, hours, and expersq. Then use black, married, hisp, and union as dummy variables, as well as others of your choice.

(Alternatively, you can use union as the target/dependent variable if you want to specify a logit or probit model.) This guide continues using lwage as the target/dependent variable.



After running the model, expand the Regression Summary panel to view coefficient estimates, goodness-of-fit statistics, and diagnostic indicators. The most important part of the output is the forecast model in the Model Results section.



The out-of-sample forecasting model for lwage with t-statistics in parentheses.

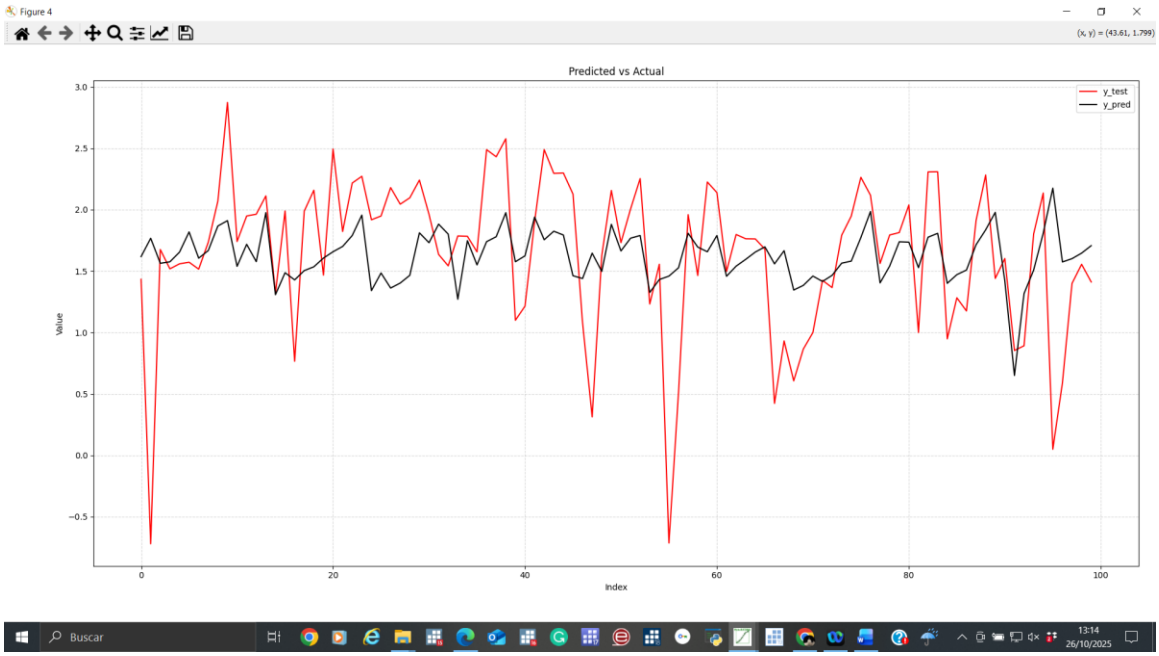
$$
\begin{aligned}
\widehat{lwage} = {} & 0.9304 \quad (15.14) \\
& + 0.0036 \cdot (exper \times educ) \quad (11.97) \\
& + 0.0037 \cdot educ^{2} \quad (13.84) \\
& + 0.1249 \cdot married \quad (7.17) \\
& + 0.0148 \cdot (union \times educ) \quad (8.96) \\
& + 5.1828 \cdot (exper^{-1} \times educ^{-1}) \quad (2.96) \\
& - 0.2381 \cdot (expersq^{-1} \times expersq^{-1}) \quad (-3.02) \\
& - 0.1236 \cdot black \quad (-4.66) \\
& - 141.6615 \cdot hours^{-1} \quad (-4.03) \\
& - 2.499 \times 10^{-8} \cdot hours^{2} \quad (-6.78)
\end{aligned}
$$

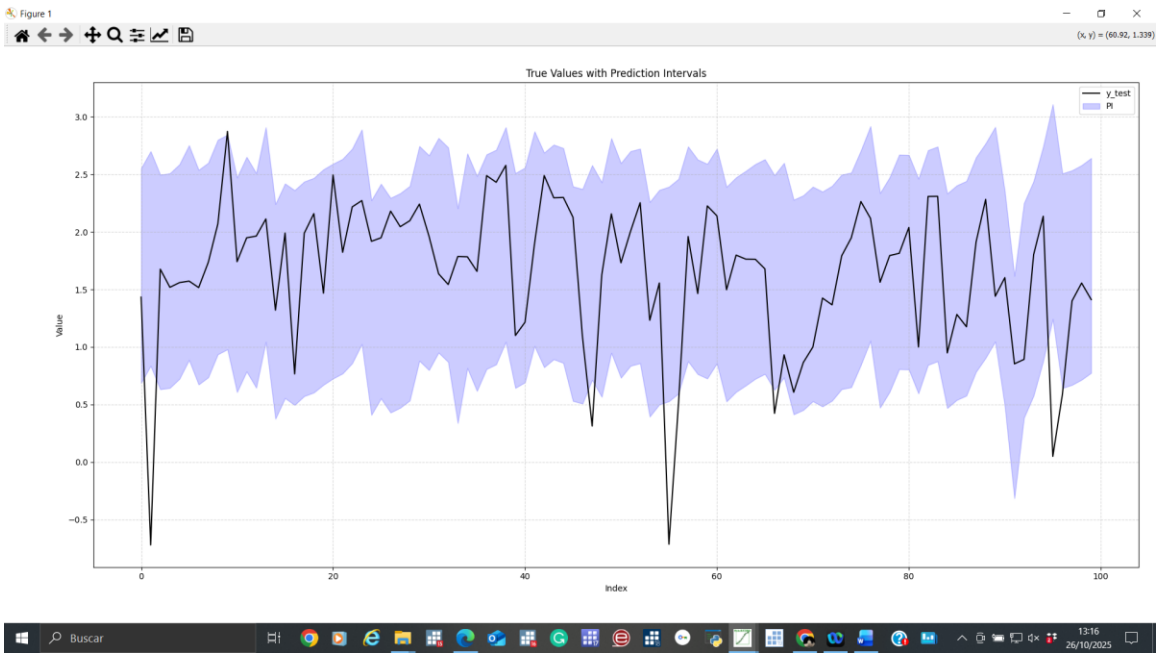
ChatGPT writes the above equation for you in a few seconds, so that you can insert it straight into your paper.

10. Out-of-Sample Forecasting

To perform an out-of-sample conditional forecast, use the

1. **Test file**: select a file to predict the dependent variable conditional on the values of the inputs. Here, use the wagepantest file with the same variable structure as the training dataset wagepantrain.
2. In **Model predictions**: you can use the standard 95% confidence interval width or choose a different one.
3. Then click **Run Predictions** to produce out-of-sample forecasts and confidence intervals.
4. **View Metrics:** temporarily disabled.
5. **Save Results:** saves the numerical predictions and intervals in a file to the current folder.
6. **Predictions and Intervals**. First, choose the number of predictions and confidence intervals that you want to visualize. Then view graphs of the predictions, actual values, and confidence intervals.





11. Summary of Capabilities

With Eco-RETINA, you can:

- Build interpretable, high-performance predictive models
- Obtain parsimonious models free of Miller's critique of in-sample bias.
- Obtain point and interval estimates with post-estimation diagnostics
- Obtain out-of-sample forecasts with forecast intervals and errors
- Compare results with OLS, Logit, Probit, panel data models, and AI algorithms

References

Capilla, J., Alcaraz, A., Valarezo, Á., García-Hiernaux, A., & Pérez Amaral, T. (2025). Eco-RETINA: A Green Flexible Algorithm for Model Building. ICAE Working Paper Series, 2501. Universidad Complutense de Madrid.

Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach* (7th ed.). Boston, MA: Cengage Learning.

APPENDIX

Description of the hyperparameters

1. **Loss**, choice of the loss function: mse, mape, aic, bic, or mae.
2. **Regression type**, you can choose ols, probit, or logit.
3. **Cross dummy**, choose to use cross products with the dummies or not.
4. **Cov type,** choose between standard covariance matrix estimates or several heteroskedasticity-consistent covariance matrix estimators.

5. **Params,** choice of transformations of the original inputs.
6. **Max R^2,** maximum coefficient of determination allowed.
7. **Grid**: grid width for the lambda parameter that controls for multicollinearity.
8. **Seed**: seed of the random number generator for generating the subsamples.
9. **Max instances**: maximum number of observations allowed.
10. **Max regressors**: maximum number of original inputs.
11. **Chunk size**: maximum number of observations that are handled at the same time.
12. **Model step**: