

Automated Residual Plot Assessment with the R Package `autovi` and the Shiny App `autovi.web`

Response to reviewers

2025-05-14

We thank the reviewers for their comments (colored in **purple** below). We have addressed it as below (in black). We also submit the difference between the previously submitted document and the revised version produced via `latexdiff`.

Referee 1

1. The Abstract could be improved further. Certain sentences in the Abstract contain grammatical errors and are hard to read.

We have revised the abstract to improve its clarity and readability. The new version is as follows.

Visual assessment of residual plots is a common approach for diagnosing linear models, but it relies on manual evaluation, which does not scale well and can lead to inconsistent decisions across analysts. The lineup protocol, which embeds the observed plot among null plots, can reduce subjectivity but requires even more human effort. In today’s data-driven world, such tasks are well-suited for automation. We present a new R package that uses a computer vision model to automate the evaluation of residual plots. An accompanying Shiny app is provided for ease of use. Given a sample of residuals, the model predicts a visual signal strength (VSS) and offers supporting information to help analysts assess the adequacy of their model fit.

2. The “`autovi.web`” RShiny app (e.g., https://akdd4t-patrick-li.shinyapps.io/autovi_web/) lacks clear instructions regarding the format of CSV files it accepts. As a result, some users may upload the original dataset instead of a two-column CSV file with fitted values and residuals.

We have updated the app to include clearer instructions on the required input format.

3. One suggestion is to consider integrating the widely-used regression model fitting functions (e.g., `lm()`, `aov()` or `glm()`) into “autovi” to automate the model fitting and assessment of residuals.

Thank you for the suggestion. We agree that this is a valuable direction for future development. However, it lies beyond the scope of the current manuscript. We have added a new Section 4.6 that discusses how the workflow could be extended to accommodate broader model classes.

4. Finally, please check the English in the manuscript. Certain lengthy sentences can be hard to understand.

We have revised the manuscript to improve clarity and grammar. Notable changes include:

- Section 2
 - Original: “VSS is an estimate of the distance between the residual distribution of a fitted classical normal linear regression model and a reference distribution; more details can be found in Li et al. 2024b.”
 - Revised: “VSS estimates the distance between the residual distribution of a fitted classical normal linear regression model and a reference distribution (see Li et al. 2024b, for details).”
- Section 5.1
 - Original: “Due to the resource allocation policy of shinyapps.io, the server enters a sleep mode during periods of inactivity, resulting in the clearing of the local Python virtual environment. Consequently, when the application “wakes up” for a new user session, these libraries need to be reinstalled.”
 - Revised: “Due to shinyapps.io’s resource policy, inactive servers enter sleep mode, clearing the local Python environment. When reactivated for a new session, libraries must be reinstalled.”
- Section 4.1
 - Original: “The third example is generated using the `surreal` package (Balamuta 2024) where structured residuals are hidden in data, to be revealed if the correct model is specified. Here a quote based on Tukey is used as the residual structure ‘visual summaries focus on unexpected values’.”
 - Revised: “The third example is generated using the `surreal` package (Balamuta 2024), where structure residuals are embedded in the data. In this case, a quote inspired by Tukey ‘visual summaries focus on unexpected values’, is used to define the residual structure.”

- Section 4.1

- Original: “This structure is blindingly obvious visually,”
- Revised: “Visually, the structure is strikingly clear,”

Referee 2

- some introduction on generating the null plot and bootstrap residual, which provides some background to understand the code summary in page 7 and Figure 3 on the comparison between VSS from the two scenarios.

We have added a new Section 3 to introduce the definitions and generation procedures for null plots and bootstrapped residuals, providing necessary context for Figure 3.

- information on the type/ scope of regression models used to train the cv model in Section 2 to understand the applicability of the cv model.

We have expanded the final paragraph of Section 2 to describe the types of regression models used for training. This provides better insight into the model’s scope and applicability.

- a brief description of each module in Figure 4, for example, what does auxiliary computation (in Figure 4) include?

We have extended the last paragraph of Section 4.5 to include brief descriptions of each module in Figure 4, including the tasks performed under “auxiliary computation”.

- the necessary changes need to make to extend the workflow to diagnose other models e.g. GLM, in Section 3.5.

This will require training and testing the computer vision model across a broader range of models, including those with different types of assumption departures. Future research could build on our prototype to support this broader class. In the meantime, we have added Section 4.6 to illustrate how the workflow could potentially be extended to other model types.

::: ::: reviewer

- the term “attention map” used in line 331 and Figure 7 caption is not introduced in the paper.

:::

We have added a definition for “attention map” as follows:

The attention map is the gradient of the model output with respect to the grayscale image input, indicating the sensitivity of the output to each pixel.

- The authors may consider moving Section 3.5 earlier in Section 3 to provide an overview of the package.

Thank you for the suggestion. We considered this structure but chose to prioritize a streamlined usage summary early in the manuscript, based on the assumption that many readers will be primarily interested in applying the tool.

- The infrastructure is built on the bandicoot OOP system, could the authors comment on this choice compared to the latest S7?

The `bandicoot` system is a custom object-oriented system tailored for our development needs, in a similar spirit to how `ggproto` is used within `ggplot2`. A detailed comparison with newer systems like S7 is beyond the scope of this paper.

- The word “be” in line 60 is redundant in the sentence.

Thank you. This has been corrected.