

## ARTICLE TEMPLATE

# Automated reading of residual plots with computer vision models

Weihao Li<sup>a</sup>

<sup>a</sup>Department of Econometrics and Business Statistics, Monash University, Clayton, VIC, Australia

### ARTICLE HISTORY

Compiled November 21, 2023

### ABSTRACT

TBD.

### KEYWORDS

TBD

## 1. Introduction

Residuals, within regression analysis, represent the differences between fitted values and observed data points, capturing the unexplained elements in the regression model. The practice of plotting residuals, advocated by influential regression literature (Cook and Weisberg 1982; Draper and Smith 1998; Belsley, Kuh, and Welsch 1980; Montgomery, Peck, and Vining 1982), serves as a standard procedure in regression diagnostics. This visual examination is crucial for identifying deviations from the model assumptions like linearity, homoscedasticity, and normality.

Generating a residual plot in most statistical software is often as straightforward as executing a line of code or clicking a button. However, accurately interpreting a residual plot can be challenging. Consider Figure 1 as an example, the residuals display a triangular shape pointing to the left. While this might suggest heteroskedasticity,

it’s important to avoid over-interpreting the visual pattern. In this case, the fitted model is correctly specified, and the triangular shape is actually a result of the skewed distribution of the predictors, rather than indicating a flaw in the model.

A residual plot can exhibit various visual features, but it’s crucial to recognize that some may arise from the characteristics of predictors and the inherent randomness of the error, rather than indicating a violation of model assumptions (Li et al. 2023). The concept of visual inference, as proposed by Buja et al. (2009), provides an inferential framework to assess whether residual plots indeed contain visual patterns inconsistent with the model assumptions. The fundamental idea involves testing whether the actual residual plot significantly differs visually from null plots, which are created using residuals generated from the null distribution. Typically, this is accomplished through the lineup protocol. In this approach, the real residual plot is embedded within a lineup alongside several null plots. If the real residual plot can be distinguished from the lineup, it provides evidence for rejecting the null hypothesis.

The practice of delivering a residual plot as a lineup is generally regarded as a valuable approach. Beyond its application in residual diagnostics, the lineup protocol has integrated into the analysis of diverse subjects. For instance, Loy and Hofmann (2013, 2014, 2015) illustrate its applicability in diagnosing hierarchical linear models. Additionally, Widen et al. (2016) demonstrates its utility in geographical research, while Krishnan and Hofmann (2021) explores its effectiveness in forensic examinations.

However, as pointed out by Li et al. (2023), a primary limitation of the lineup protocol lies in its reliance on human judgments. Unlike conventional statistical tests that can be performed numerically and automatically in statistical software, the lineup protocol requires human evaluation of images. This characteristic makes it less suitable for large-scale applications, given the associated high labour costs and time requirements.

There is a compelling need to develop an approach that alleviates people’s workload by automating repetitive tasks and providing standardized results in a controlled environment. The large-scale evaluation of lineups is impractical without the use of technology and machines.

The utilization of computers to interpret data plots has a rich history, with early efforts such as “Scagnostics” by Tukey and Tukey (1985), focusing on scatterplot

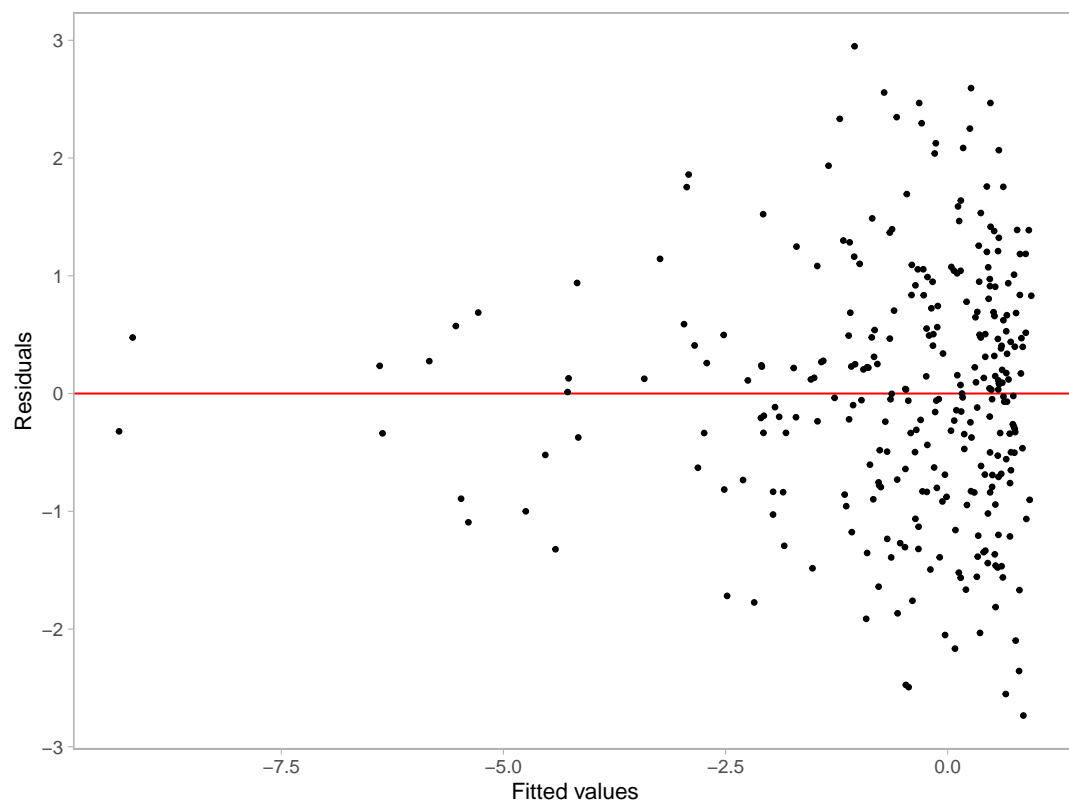
diagnostics. Wilkinson, Anand, and Grossman (2005) expanded on this work in 2005, introducing graph theoretic scagnostics, which encompassed nine computable measures applied to planar proximity graphs. These measures, including “Outlying,” “Skinny,” “Stringy,” “Straight,” “Monotonic,” “Skewed,” “Clumpy,” and “Striated,” aimed to characterize outliers, shape, density, trend, and coherence of the data.

While this approach has been inspiring, there is a recognition that it may not capture all the necessary visual features distinguishing actual residual plots from null plots. An alternative, more promising approach involves allowing computers to learn the visual feature extraction function. In essence, this means enabling computers to discern which visual features are significant for residual diagnostics and how to extract them. Modern computer vision models, a subset of artificial intelligence, present a compelling solution to this challenge.

Modern computer vision models often rely on deep neural networks with convolutional layers (Fukushima and Miyake 1982). These layers leverage hierarchical patterns in data, downsizing and transforming images by summarizing information in a small space. Numerous studies have demonstrated the efficacy of convolutional layers in addressing various vision tasks, including image recognition (Rawat and Wang 2017).

Despite the widespread use of computer vision models in fields like computer-aided diagnosis (Lee and Chen 2015), pedestrian detection (Brunetti et al. 2018), and facial recognition (Emami and Suciú 2012), their application in reading data plots remains limited. While some studies have explored the use of computer vision models for tasks such as reading recurrence plots for time series regression (Ojeda, Solano, and Peramo 2020), time series classification (Chu et al. 2019; Hailesilassie 2019; Hatami, Gavet, and Debayle 2018; Zhang et al. 2020), anomaly detection (Chen, Su, and Yang 2020), and pairwise causality analysis (Singh et al. 2017), the application of reading residual plots with computer vision models represents a relatively new field of study.

In this paper, we develop computer vision models and integrate them into the residual plots diagnostics workflow, filling the gap of . . . The paper is structured as follows: . . .



**Figure 1.** An example residual vs fitted values plot (red line indicates 0). The vertical spread of the data points varies with the fitted values. This often indicates the existence of heteroskedasticity.

## 2. Methodology

### 2.1. *Generation of simulated training data*

### 2.2. *Architecture of the computer vision model*

### 2.3. *Training process and hyperparameter tuning*

## 3. Results

### 3.1. *Model evaluation*

- Metrics for model performance
- Shap values
- Heatmap

### 3.2. *Comparison with human visual inference*

#### 3.2.1. *Overview of the human subject experiment*

#### 3.2.2. *Comparison*

- power comparison
- decisions

### 3.3. *Case study 1: ...*

### 3.4. *Case study 2: ...*

### 3.5. *Case study 3: datasaurus*

## 4. Discussion

## 5. Conclusion

- Summary of findings
- Contributions to the field
- Future directions for research

## References

- Belsley, David A, Edwin Kuh, and Roy E Welsch. 1980. *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.
- Brunetti, Antonio, Domenico Buongiorno, Gianpaolo Francesco Trotta, and Vitoantonio Bevilacqua. 2018. “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey.” *Neurocomputing* 300: 17–33.
- Buja, Andreas, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun-Kyung Lee, Deborah F Swayne, and Hadley Wickham. 2009. “Statistical inference for exploratory data analysis and model diagnostics.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361–4383.
- Chen, Yun, Shijie Su, and Hui Yang. 2020. “Convolutional neural network analysis of recurrence plots for anomaly detection.” *International Journal of Bifurcation and Chaos* 30 (01): 2050002.
- Chu, Hongyang, Xinwei Liao, Peng Dong, Zhiming Chen, Xiaoliang Zhao, and Jiandong Zou. 2019. “An automatic classification method of well testing plot based on convolutional neural network (CNN).” *Energies* 12 (15): 2846.
- Cook, R Dennis, and Sanford Weisberg. 1982. *Residuals and influence in regression*. New York: Chapman and Hall.
- Draper, Norman R, and Harry Smith. 1998. *Applied regression analysis*. Vol. 326. John Wiley & Sons.
- Emami, Shervin, and Valentin Petrut Suci. 2012. “Facial recognition using OpenCV.” *Journal of Mobile, Embedded and Distributed Systems* 4 (1): 38–43.
- Fukushima, Kunihiro, and Sei Miyake. 1982. “Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position.” *Pattern recognition* 15 (6): 455–469.
- Hailesilassie, Tameru. 2019. “Financial Market Prediction Using Recurrence Plot and Convolutional Neural Network.” .
- Hatami, Nima, Yann Gavet, and Johan Debayle. 2018. “Classification of time-series images using deep convolutional neural networks.” In *Tenth international conference on machine vision (ICMV 2017)*, Vol. 10696, 242–249. SPIE.
- Krishnan, Ganesh, and Heike Hofmann. 2021. “Hierarchical Decision Ensembles-An inferential framework for uncertain Human-AI collaboration in forensic examinations.” *arXiv preprint arXiv:2111.01131* .

- Lee, Howard, and Yi-Ping Phoebe Chen. 2015. “Image based computer aided diagnosis system for cancer detection.” *Expert Systems with Applications* 42 (12): 5356–5365.
- Li, Weihao, Dianne Cook, Emi Tanaka, and Susan VanderPlas. 2023. “A Plot is Worth a Thousand Tests: Assessing Residual Diagnostics with the Lineup Protocol.” *arXiv preprint arXiv:2308.05964* .
- Loy, Adam, and Heike Hofmann. 2013. “Diagnostic tools for hierarchical linear models.” *Wiley Interdisciplinary Reviews: Computational Statistics* 5 (1): 48–61.
- Loy, Adam, and Heike Hofmann. 2014. “HLMdiag: A suite of diagnostics for hierarchical linear models in R.” *Journal of Statistical Software* 56: 1–28.
- Loy, Adam, and Heike Hofmann. 2015. “Are you normal? The problem of confounded residual structures in hierarchical linear models.” *Journal of Computational and Graphical Statistics* 24 (4): 1191–1209.
- Montgomery, Douglas C, Elizabeth A Peck, and G Geoffrey Vining. 1982. *Introduction to linear regression analysis*. John Wiley & Sons.
- Ojeda, Sun Arthur A, Geoffrey A Solano, and Elmer C Peramo. 2020. “Multivariate time series imaging for short-term precipitation forecasting using convolutional neural networks.” In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, 33–38. IEEE.
- Rawat, Waseem, and Zenghui Wang. 2017. “Deep convolutional neural networks for image classification: A comprehensive review.” *Neural computation* 29 (9): 2352–2449.
- Singh, Karamjit, Garima Gupta, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal. 2017. “Deep convolutional neural networks for pairwise causality.” *arXiv preprint arXiv:1701.00597* .
- Tukey, John W, and Paul A Tukey. 1985. “Computer graphics and exploratory data analysis: An introduction.” In *Proceedings of the sixth annual conference and exposition: computer graphics*, Vol. 85, 773–785.
- Widen, Holly M, James B Elsner, Stephanie Pau, and Christopher K Uejio. 2016. “Graphical inference in geographical research.” *Geographical Analysis* 48 (2): 115–131.
- Wilkinson, Leland, Anushka Anand, and Robert Grossman. 2005. “Graph-theoretic scagnostics.” In *Information Visualization, IEEE Symposium on*, 21–21. IEEE Computer Society.
- Zhang, Ye, Yi Hou, Shilin Zhou, and Kewei Ouyang. 2020. “Encoding time series as multi-scale signed recurrence plots for classification using fully convolutional networks.” *Sensors* 20 (14): 3818.