The Journal of Open Source Software

lpcde: Estimation and Inference for Local Polynomial Conditional Density Estimators

Matias D. Cattaneo [1], **Rajita Chandak** [2¶], **Michael Jansson** [3], and **Xinwei Ma** [4]

1 Department of Operations Research and Financial Engineering, Princeton University, USA **2** Institute of Mathematics, EPFL, Switzerland **3** Department of Economics, University of California, Berkeley, USA **4** Department of Economics, University of California, San Diego, USA ¶ Corresponding author

DOI: 10.21105/joss.07241

Software

- Review
- Repository
- Archive

Editor: Susan Holmes

Reviewers:

- @salbalkus
- @soichiroy

Submitted: 26 August 2024
Published: 07 March 2025

Summary

Conditional cumulative distribution functions (CDFs), conditional probability density functions (PDFs), and derivatives thereof, are important parameters of interest in statistics, econometrics, and other data science disciplines. The package `lpcde` implements new estimation and inference methods for conditional CDFs, conditional PDFs, and derivatives thereof, employing the kernel-based local polynomial smoothing approach introduced in Cattaneo et al. (2024a).

The package `lpcde` offers data-driven (pointwise and uniform) estimation and inference methods for conditional CDFs, conditional PDFs, and derivatives thereof, which are automatically valid at both interior and boundary points of the support of the outcome and conditioning variables. For point estimation, the package offers mean squared error optimal bandwidth selection and associated optimal mean square and uniform point estimators. For inference, the package offers valid confidence intervals and confidence bands based on robust bias-correction techniques (Calonico et al., 2018, 2022). Finally, these statistical procedures can be easily used for visualization and graphical presentation of smooth estimates of conditional CDFs, conditional PDFs, and derivative thereof, with custom `ggplot` (Wickham, 2016) commands built for the package.

This package is currently the only open source implementation of an estimator offering boundary adaptive, data-driven conditional density estimation with robust bias-corrected pointwise confidence interval and uniform confidence band constructions, providing users with statistical tools to better understand the reliability of their empirical analysis. A detailed tutorial, replication files, and other information on how to use the package can be found in the GitHub repository and through the CRAN repository. See also the `lpcde` package website (https://nppackages.github.io/lpcde/) and the companion arXiv article (Cattaneo et al., 2024b) for additional methodological information and numerical results.

Statement of need

Wand & Jones (1995), Fan & Gijbels (1996), Simonoff (2012), and Scott (2015) give textbook introductions to kernel-based density and local polynomial estimation and inference methods. The core idea underlying the estimator implemented in `lpcde` is to use kernel-based local polynomial smoothing methods to construct an automatic boundary adaptive estimator for conditional CDFs, conditional PDFs, and derivatives thereof. The estimator implemented in this package consists of two steps. The first step estimates the conditional distribution function using standard local polynomial regression methods, and the second step applies local polynomial smoothing to the (non-smooth) local polynomial conditional CDF estimate from the first step to obtain a smooth estimate of the conditional CDF, conditional PDF, and derivatives thereof.

A distinct advantage of this estimation method over existing ones is its boundary adaptivity for a possibly unknown compact support of the data. Furthermore, the estimator has a simple closed form representation, which leads to easy and fast implementation. Unlike other boundary adaptive procedures, the estimation procedures implemented in the package `lpcde` do not require pre-processing of data, and thus avoid the challenges of hyper-parameter tuning: only one bandwidth parameter needs to be selected for implementation. See Cattaneo et al. (2024a) and Cattaneo et al. (2024b) for more details.

Comparing and contrasting existing toolsets

The package `lpcde` contributes to a small set of open source statistical software packages implementing estimation and inference methods for conditional CDF, conditional PDF, and derivatives thereof. More specifically, we identified three `R` packages, `hdrcde` (Hyndman et al., 2021), `haldensify` (Hejazi et al., 2022), and `np` (Hayfield & Racine, 2008), and one `Python` package, `cde` (Rothfuss et al., 2019), which provide related methodology. There are no open source `Stata` packages that implement comparable estimation and inference methods. The table below summarizes some of the main differences between those other packages and `lpcde`. Notably, `lpcde` is the only package available that provides both pointwise and uniform uncertainty quantification, in addition to producing boundary adaptive mean square and uniformly optimal point estimates via data-driven, optimal tuning parameter selection. Furthermore, the `lpcde` package produces proper conditional density estimates that are non-negative and integrate to one. These features are unique contributions of the package to the `R` toolkit and, more broadly, to the open source statistical community.

Package	Programming language	CDF/Derivative estimation	Regularized density	Valid at boundary	Standard error	Valid inference	Confidence bands	Bandwidth selection
`hdrcde`	R	×	×	×	×	×	×	✓
`np`	R	×	×	×	✓	×	×	✓
`haldensify`	R	×	×	×	✓	×	×	✓
`cde`	Python	×	×	×	×	×	×	✓
`lpcde`	R	✓	✓	✓	✓	✓	✓	✓

Acknowledgements

Cattaneo gratefully acknowledges financial support from the National Science Foundation through grants SES-1947805, DMS-2210561, and SES-2241575, and from the National Institute of Health (R01 GM072611-16). Jansson gratefully acknowledges financial support from the National Science Foundation through grant SES-1947662.

References

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, *113*(522), 767–779. https://doi.org/10.1080/01621459.2017.1285776

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2022). Coverage error optimal confidence intervals for local polynomial regression. *Bernoulli*, *28*(4), 2998–3022. https://doi.org/10.3150/21-bej1445

Cattaneo, M. D., Chandak, R., Jansson, M., & Ma, X. (2024a). Boundary adaptive local polynomial conditional density estimators. *Bernoulli*, *30*(4), 3193–3223. https://doi.org/10.3150/23-BEJ1711

Cattaneo, M. D., Chandak, R., Jansson, M., & Ma, X. (2024b). Lpcde: Estimation and inference for local polynomial conditional density estimators. *arXiv Preprint arXiv:2204.10375*. https://arxiv.org/abs/2204.10375

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall/CRC. https://doi.org/10.1201/9780203748725

Hayfield, T., & Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, *27*(5). https://doi.org/10.18637/jss.v027.i05

Hejazi, N. S., Laan, M. J. van der, & Benkeser, D. (2022). 'Haldensify': Highly adaptive lasso conditional density estimation in 'R'. *Journal of Open Source Software*, *7*(77), 4522. https://doi.org/10.21105/joss.04522

Hyndman, R. J., Einbeck, J., & Wand, M. P. (2021). *Hdrcde: Highest density regions and conditional density estimation*. https://doi.org/10.32614/cran.package.hdrcde

Rothfuss, J., Ferreira, F., Walther, S., & Ulrich, M. (2019). Conditional density estimation with neural networks: Best practices and benchmarks. *arXiv:1903.00954*.

Scott, D. W. (2015). *Multivariate density estimation: Theory, practice, and visualization*. John Wiley & Sons. https://doi.org/10.2307/1270280

Simonoff, J. S. (2012). *Smoothing methods in statistics*. Springer–Verlag. https://doi.org/10.1007/978-1-4612-4026-6

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*. Chapman & Hall/CRC. https://doi.org/10.1002/9781118890370

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://doi.org/10.1007/978-0-387-98141-3