

# bunching: A package for implementing the bunching estimator in R

24 January 2022

## Summary

(Clifford and Mavrokonstantis 2021) The popularity of the bunching estimator has surged in recent years, particularly in the field of social science.

The bunching estimator has experienced a growing popularity in recent years, especially in social science applications. This article introduces the bunching package, an R implementation of the estimator which enables the user to conduct such bunching analysis in a kink or notch setting. It begins by outlining the optimization theory underlying the bunching estimator. It then introduces the package’s main functions and provides a tutorial on how to implement the estimator in R using bunching, with many examples of different cases of bunching at kinks and notches. Finally, a technical section on the estimation methods underlying the estimator points out several nuances for users interested in the implementation details.

Recent years have seen a growing adoption of the bunching estimator, especially in social science applications. While the original aim of the estimator was to tackle questions within the fields of labor and public economics (and taxation in particular), it can be applied to any setting where a constrained optimization problem involves a discontinuity in a constraint, which can be related to a discontinuity in the observed density of the decision variable. It naturally extends to any setting that involves prices, since taxes are simply instruments that scale or shift prices. Indeed, it has since been applied in numerous studies outside of economics, including empirical questions in psychology (Allen et al. 2017), law (Dharmapala 2018) and finance (Dagostino 2019). Due to its generality, it can be readily used in any commercial setting involving non-linearities in constraints, such as non-linear pricing schemes, non-linear incentive contracts in the workplace, and so on.

This software has enabled recent scholarly publications (Alberini and Ščasnýc 2022; Clifford and Mavrokonstantis 2021; Mavrokonstantis and Seibold 2022), as well as ongoing research [TO ADD].

The bunching package provides an R implementation of the bunching estimator, as developed (in different flavors and applications) by Saez (2010), Chetty et al. (2011) and Kleven and Waseem (2013). The goal of the estimator is to credibly estimate the resulting bunching at a particular level of a choice set that features a discontinuity, i.e. the (excess) mass over and above what would have been observed in the absence of the discontinuity in the given constraint. The estimator solves a challenge of both causal inference and prediction, since we would like to (causally) attribute the excess mass to a particular discontinuity in a constraint, but cannot typically observe the counterfactual density in the absence of that constraint. The estimator’s novelty is to provide a prediction method for this counterfactual by combining reduced-form methods with micro-economically founded optimization theory, while relying on minimal assumptions. Two main types of discontinuities are considered: kinks, i.e. changes in the slopes of choice sets, and notches, i.e. changes in the levels of choice sets. This paper presents the details behind the R package bunching, which enables the user to conduct such bunching analysis in a kink or notch setting, and returns a rich set of results. Important features of the package include functionality to control for (different levels of) round-number bunching or other bunching masses within the estimation bandwidth, options to split bins by placing the bunching point as the minimum, median or maximum in its bin (for robustness analysis), and can return estimates of both parametric and reduced-form versions of elasticities associated with the bunching mass. It also provides an exploratory visualization function to speed up pre-analysis without having to run any actual estimations. A bunching plot is a very important output in every bunching analysis, since bunching estimates must always be accompanied by compelling graphical evidence. To this end, bunching produces plots in the standardized expected style since Chetty et al. (2011), and offers the user a plethora of options to customize the plot’s

appearance. Further, it returns bootstrapped estimates of all the main estimable parameters, which can be used for further analysis such as incorporation into structural models that rely on bunching moments.

## Statement of Need

To my knowledge, the only other implementation of the bunching estimator in R is provided by the package `bunchr` (Trilnick 2017). While this provides some useful functionality for bunching analysis, it has several limitations, which the bunching package overcomes. For instance, `bunchr` does not allow the user to control for round-number bunching or for the presence of other bunching masses within the estimation bandwidth. While these cases are not discussed in the textbook version of bunching theory, they are usually present in most empirical settings.<sup>1</sup> The inability to control for such effects can lead to highly biased estimates because the presence of additional bunching masses besides the one of interest will have a large impact on the estimated counterfactual density. Another limitation is that `bunchr` implements case-resampling, instead of the estimator’s required residual-resampling to estimate standard errors of the bunching mass, and it does not return the vector of bootstrapped estimates. This is a significant drawback because bunching estimates are often incorporated into structural models which require the full set of bootstrapped values for estimation of standard errors of structural parameters.<sup>2</sup> In addition, it is limited in the estimated moments it returns, with some crucial ones missing. For instance, it returns the excess mass but not the normalized estimate of this (or equivalently, the estimated height of the counterfactual at the bunching point). Without this, it is impossible to compare bunching masses at different levels which may feature different counterfactual levels. All of the aforementioned features missing from `bunchr` are supported by the bunching package. One other important difference is that bunching creates plots in a different style, which has been designed to match the standardized format found in published articles since Chetty et al. (2011). It also provides much more functionality to control for a wide range of graphical parameters (such as line, marker, axes styles, etc.) to customize the final plot output. The remainder of this article presents the details of the package `bunching` (Mavrokonstantis 2019), which implements the bunching estimator in R and is available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=bunching>. Section 2 is meant to be a beginners’ introduction to the optimization theory underlying the bunching estimator. Section 3 provides an overview of the package and discusses its two main functions, `bunchit()` and `plot_hist()`, and Section 4 provides a tutorial on how to implement the estimator in R with `bunching`, with many examples of different cases of bunching at kinks and notches. Section 5 concludes with a discussion of future avenues for the package. The article also contains a technical appendix which details the estimation technique behind the estimator and points out some associated nuances.

The forces on stars, galaxies, and dark matter under external gravitational fields lead to the dynamical evolution of structures in the universe. The orbits of these bodies are therefore key to understanding the formation, history, and future state of galaxies. The field of “galactic dynamics,” which aims to model the gravitating components of galaxies to study their structure and evolution, is now well-established, commonly taught, and frequently used in astronomy. Aside from toy problems and demonstrations, the majority of problems require efficient numerical tools, many of which require the same base code (e.g., for performing numerical orbit integration).

## Statement of need

`Gala` is an Astropy-affiliated Python package for galactic dynamics. Python enables wrapping low-level languages (e.g., C) for speed without losing flexibility or ease-of-use in the user-interface. The API for `Gala` was designed to provide a class-based and user-friendly interface to fast (C or Cython-optimized) implementations of common operations such as gravitational potential and force evaluation, orbit integration, dynamical transformations, and chaos indicators for nonlinear dynamics. `Gala` also relies heavily on and interfaces well with the implementations of physical units and astronomical coordinate systems in the `Astropy` package (Astropy Collaboration 2013) (`astropy.units` and `astropy.coordinates`).

`Gala` was designed to be used by both astronomical researchers and by students in courses on gravitational dynamics or astronomy. It has already been used in a number of scientific publications (Pearson, Price-

Whelan, and Johnston 2017) and has also been used in graduate courses on Galactic dynamics to, e.g., provide interactive visualizations of textbook material (Binney and Tremaine 2008). The combination of speed, design, and support for Astropy functionality in **Gala** will enable exciting scientific explorations of forthcoming data releases from the *Gaia* mission (Gaia Collaboration 2016) by students and experts alike.

## Mathematics

Single dollars (\$) are required for inline mathematics e.g.  $f(x) = e^{\pi/x}$

Double dollars make self-standing equations:

$$\Theta(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{else} \end{cases}$$

You can also use plain L<sup>A</sup>T<sub>E</sub>X for equations

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x)e^{i\omega x}dx \tag{1}$$

and refer to Equation 1 from text.

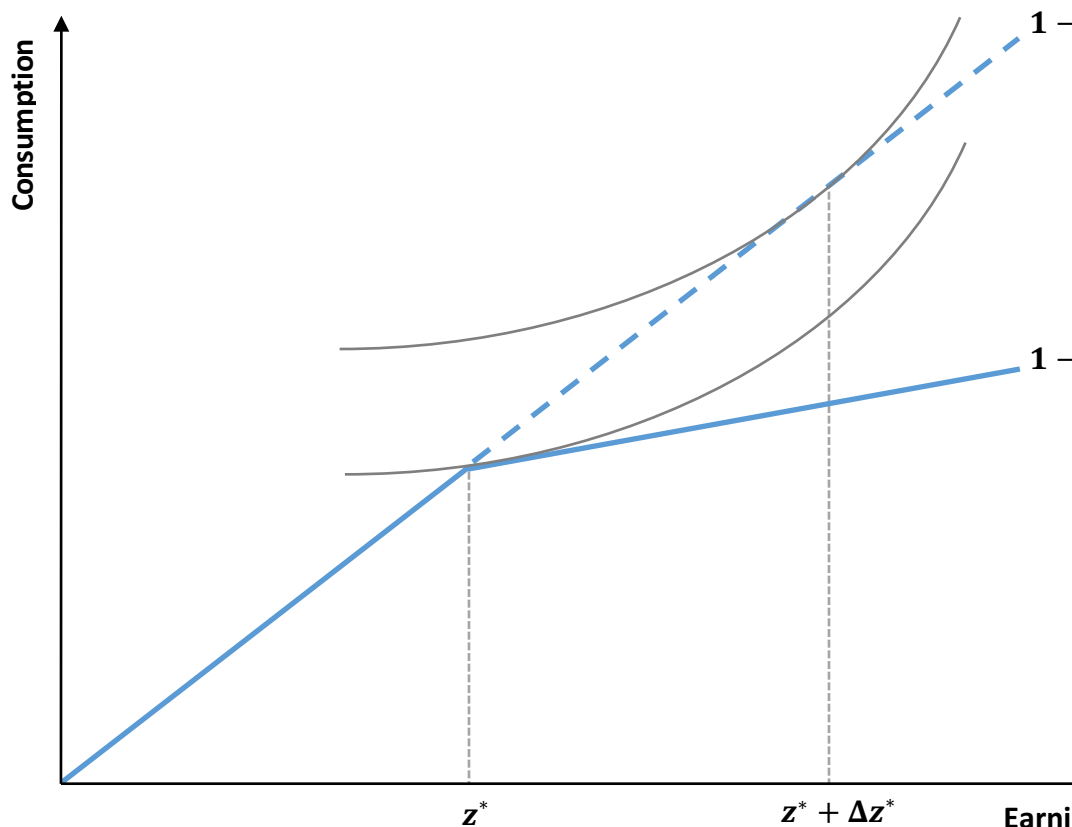
## Citations

Citations to entries in paper.bib should be in rMarkdown format.

If you want to cite a software repository URL (e.g. something on GitHub without a preferred citation) then you can do it with the example BibTeX entry below for Smith, Thaney, and Hahnel (2020).

For a quick reference, the following citation commands can be used: - @author:2001 -> “Author et al. (2001)”  
- [author:2001] -> “(Author et al., 2001)” - [author1:2001; author2:2001] -> “(Author1 et al., 2001; Author2 et al., 2002)”

## Figures



Figures can be included like this:  
and referenced from text using ??.

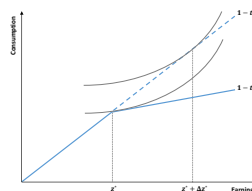


Figure sizes can be customized by adding an optional second parameter:

## Acknowledgements

We acknowledge contributions from Brigitta Sipocz, Syrtis Major, and Semyeong Oh, and support from Kathryn Johnston during the genesis of this project.

## References

- Alberini, Bezhanishvili, A., and M. Ščasnýc. 2022. “Wild tariff schemes: Evidence from the Reuplic of Georgia.” *Energy Economics*, June. <https://www.sciencedirect.com/science/article/pii/S0140988322001992>.
- Astropy Collaboration. 2013. “Astropy: A community Python package for astronomy.” *Astronomy and Astrophysics* 558 (October). <https://doi.org/10.1051/0004-6361/201322068>.
- Binney, J., and S. Tremaine. 2008. *Galactic Dynamics: Second Edition*. Princeton University Press. <http://adsabs.harvard.edu/abs/2008gady.book.....B>.
- Clifford, S., and P. Mavrokonstantis. 2021. “Tax Enforcement Using a Hybrid between Self- and Third-party Reporting.” *Journal of Public Economics*, November. <https://www.sciencedirect.com/science/article/abs/pii/S0047272721001559>.

- Dagostino, R. 2019. “The Impact of Bank Financing on Municipalities’ Bond Issuance and the Real Economy.” Brookings Institute. [https://www.brookings.edu/wp-content/uploads/2019/07/Dagostino2019\\_muni.pdf](https://www.brookings.edu/wp-content/uploads/2019/07/Dagostino2019_muni.pdf).
- Gaia Collaboration. 2016. “The Gaia mission.” *Astronomy and Astrophysics* 595 (November). <https://doi.org/10.1051/0004-6361/201629272>.
- Mavrokonstantis, P., and A. Seibold. 2022. “Bunching and Adjustment Costs: Evidence from Cypriot Tax Reforms.” *Journal of Public Economics*, October. <https://www.sciencedirect.com/science/article/abs/pii/S0047272722001293>.
- Pearson, S., A. M. Price-Whelan, and K. V. Johnston. 2017. “Gaps in Globular Cluster Streams: Pal 5 and the Galactic Bar.” *ArXiv e-Prints*, March. <http://adsabs.harvard.edu/abs/2017arXiv170304627P>.
- Smith, A. M., K. Thaney, and M. Hahnel. 2020. “Fidgit: An Ungodly Union of GitHub and Figshare.” *GitHub Repository*. GitHub. <https://github.com/arfon/fidgit>.