The Journal of Open Source Software

MEDUSA: An R tool to model equity and distributional impacts

Eva Alonso-Epelde [1] and **Clàudia Rodés-Bachs** [1]

1 Basque Centre for Climate Change (BC3), Edificio Sede 1-1, Parque Científico de UPV/EHU, Barrio Sarriena s/n, 48940 Leioa, Spain

DOI: 10.21105/joss.07697

Software

- Review
- Repository
- Archive

Editor: Nikoleta Glynatsi

Reviewers:

- @thegargiulian
- @apwheele

Submitted: 23 September 2024
Published: 07 July 2025

Summary

Addressing 21st-century challenges, such as climate change, demands policies that not only advance social justice but also prevent the exacerbation of existing inequalities. While Integrated Assessment Models (IAMs) are a fundamental tool to carry out impact analyses of policies from a holistic perspective, micro-simulation models are crucial for identifying heterogeneous socioeconomic impacts and ensuring fairer and more targeted policies.

`medusa` is an R package designed for conducting distributional analyses, either independently or in conjunction with other models, including IAMs. The extensive database in which the microsimulation model is based allows highly disaggregated results by considering a wide range of socioeconomic and demographic characteristics of households, such as income level, place of residence, type of family, and feminization degree. The package combines this detailed household data with the calculation of energy and transport poverty indices. The structure of the `medusa` package is summarized in Figure 1.

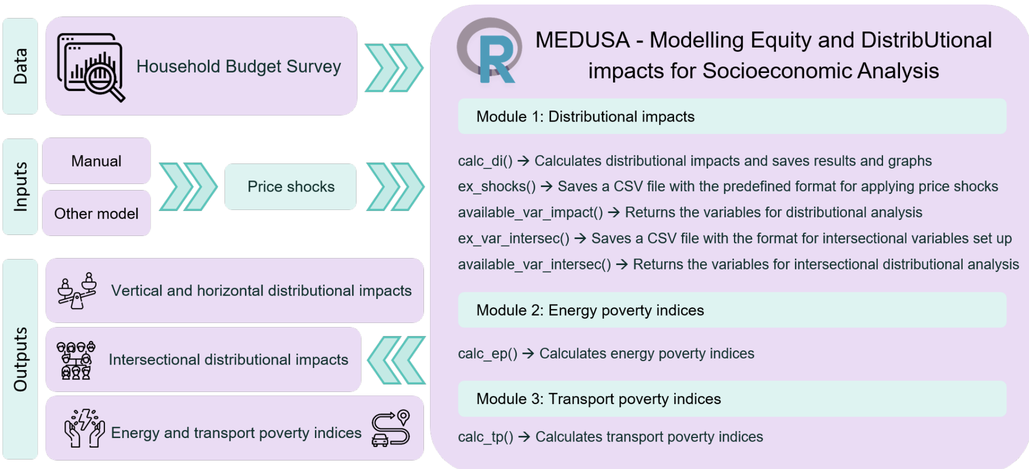

Figure 1: Structure of the `medusa` package

The `medusa` package is available online through the public domain https://github.com/bc3LC/medusa. Below is a simplified code example demonstrating how to execute the package. For a comprehensive introduction to `medusa`, a detailed step-by-step tutorial is provided in the form of an R vignette, accessible here.

```r
install.packages("remotes")
library(remotes)
install_github("bc3LC/medusa")
```

```r
library(medusa)

# Download the example file to enter price shocks
ex_shocks()

# After introducing the price shocks in the csv file, upload the edited file
exampledf <- read.csv(file_name,    # File name or full file path
              header = TRUE,        # Header indicator (set to TRUE)
              sep = ",",            # Value separator used in the file
              dec = ".",            # Decimal point format used
              ...)                  # Additional arguments to be passed

# Calculate distributional impacts
calc_di( year,                      # Base year for the simulation
         elevate = F,               # Elevation of raw data (set to FALSE)
         shocks = exampledf,        # Name of the uploaded file with shocks
         var_impact = "DECILE",     # Socioeconomic variable to be used
         ...)                       # Additional arguments to be passed
```

Statement of need

Addressing critical challenges like climate change requires ambitious policies that promote social justice without worsening existing inequalities, such as income or gender disparities (E. Alonso-Epelde et al., 2024). To ensure this, it is essential to conduct policy impact assessments that not only consider the economy, energy, land, and water systems holistically but also analyze the distributional impacts across different population groups (Bazoli et al., 2022; Walker, 2010). While Integrated Assessment Models (IAMs) (Weyant, 2017) are invaluable in policy evaluation (Van Beek et al., 2020), they often lack the granularity needed to assess socio-economic disparities. Micro-simulation models for distributional analysis fill this gap by providing detailed, heterogeneous results, enabling policymakers to identify vulnerable populations and implement targeted compensatory measures (Tomás et al., 2023). This ensures that policies are equitable and socially just.

`medusa` facilitates distributional impact analyses through an overnight-effect microsimulation model, leveraging microdata from the Household Budget Survey (HBS),[1] a standardized and comprehensive dataset available across EU countries (Eurostat, 2003). The HBS offers detailed insights into household consumption patterns and socioeconomic characteristics at both household and individual levels, allowing for highly granular analysis. This enables the integration of an intersectional approach[2] considering factors such as class, gender, and race, and provides more robust and nuanced results for assessing policy impacts on diverse population groups.

The results derived from the model are presented as the relative impact ($\%$) on total equivalent consumption expenditure.[3] The relative impact, Δe_h^s, shows the additional cost that household

[1]Since the package currently operates for Spain, the microdata used in this package come from the Spanish National Statistics Institute (INE). They can be accessed at this link.

[2]Intersectionality refers to the fact that the privileges or oppression of each individual depend on the multiple social categories to which he or she belongs, which are social constructs and can change over time (Cho et al., 2013; Crenshaw, 1994; Davis, 1983; Djoudi et al., 2016; Kaijser & Kronsell, 2014). Intersectionality is therefore also a tool for analysing the articulation of different socio-economic categories (e.g. class, gender, race, etc.) rather than considering them as independent forms of power relations (Colombo & Rebughini, 2016).

[3]Equivalent consumption expenditure is used instead of income as it is considered a better proxy for permanent household income since it fluctuates less in the long run (Goodman & Oldfield, 2004). The equivalent spending is calculated based on household spending relativized by the modified OECD equivalence scale, thus considering the economies of scale generated in households based on their size. The modified OECD scale values 1 for the reference person in the household, 0.5 for other people aged 14 or over, and 0.3 for other people under 14 years of age.

h would assume in a proposed scenario s in relative terms (%), compared to the initial household expenditure. It is calculated as:

$$\Delta e_h^s = \frac{\sum_c e_{c,h}(1+\Delta p_c^s) - \sum_c e_{c,h}}{\sum_c e_{c,h}} \times 100$$

here $e_{c,h}$ refers to the total spending on each consumption category c consumed by each household h in the baseline scenario and Δp_c^s is the increase in prices by consumption category and scenario obtained with the price model.

To the best of our knowledge, there was no existing software specifically designed for conducting distributional impact assessments of policies in an accessible and reproducible way. While some economic models allow for similar analyses, they are often highly specialized, require advanced technical expertise, or are not openly available. medusa fills this gap by providing the first R package dedicated to this type of analysis. It offers a transparent and user-friendly framework that enables a broader audience—including researchers, analysts, and policymakers—to incorporate equity considerations into policy evaluation. This democratization of access to distributional tools helps address a critical shortcoming in existing modeling approaches, supporting the design of more inclusive and just policies.

Functionality

The `medusa` package includes several functions that have been classified in 3 main modules. Note that these functions are listed in an R vignette, which includes a step-by-step tutorial.

- Module 1: Functions to calculate the distributional impacts. The main function for users, `calc_di()`, calculates the distributional impacts for different households according to a wide range of socioeconomic and demographic characteristics. The distributional impacts could be calculated for one or more intersecting variables. When introducing the outputs of a macro model in `medusa`, the microdata in which the microsimulation model is based should be elevated to the National Accounts (Cazcarro et al., 2022), this can be easily done indicating TRUE in the `elevate` parameter. Furthermore, in order to facilitate the analysis of the results, the package allows the generation of summary dataframes and figures of the distributional impacts either for one or several socioeconomic variables.
- Module 2: Functions to calculate energy poverty indices.The main function for users, `calc_ep()`, calculates the energy poverty index for the selected year/s and the selected indicator. The indicators included in the package are the 10%, 2M, LIHC, HEP and HEP_LI. These indicators have been commonly used in the literature to measure energy poverty during the last decades and are explained here.
- Module 3: Functions to calculate transport poverty indices.The main function for users, `calc_tp()`, calculates the transport poverty index for the selected year/s and the selected indicator. The indicators included in the package are the 10%, 2M, LIHC and VTU. These indicators are based on the proposal by E. Alonso-Epelde et al. (2023) and are explained here.

The package includes default input files (.Rda), which are required for running the various functions, simplifying the process for users.

Output files are generated in both comma-separated values (CSV) and portable network graphics (PNG) formats, with user control over file creation. When the save parameter is set to TRUE, the function saves a CSV file containing the selected results in the defined directory. Additionally, if the `fig` parameter is set to TRUE, the function produces and saves a bar plot to visualize the corresponding output.

The package is actively evolving to meet research and policy needs, with several new features planned for future releases. For instance, an upcoming update will extend simulation capabilities to all EU countries, as the initial release currently covers only Spain. Additionally, we are developing a user interface designed to enable individuals without R programming expertise to perform socioeconomic analyses effectively.

Acknowledgements

E.A. and C.R. acknowledges financial support from the European Union's Horizon research program under grant agreement 101056306 (IAM COMPACT project).

References

Alonso-Epelde, E., García-Muros, X., & González-Eguino, M. (2024). Climate action from a gender perspective: A systematic review of the impact of climate policies on inequality. *Energy Research & Social Science*, *112*, 103511. https://doi.org/10.1016/j.erss.2024.103511

Alonso-Epelde, E., García-Muros, X., & González-Eguino, M. (2023). Transport poverty indicators: A new framework based on the household budget survey. *Energy Policy*, *181*, 113692. https://doi.org/10.1016/j.enpol.2023.113692

Bazoli, N., Burlacu, S., Fiorio, C., Marzadro, S., Pycroft, J., Vergolini, L., & others. (2022). *Study on distributional impact assessment*. https://doi.org/10.2767/511644

Cazcarro, I., Amores, A. F., Arto, I., & Kratena, K. (2022). Linking multisectoral economic models and consumption surveys for the european union. *Economic Systems Research*, *34*(1), 22–40. https://doi.org/10.1080/09535314.2020.1856044

Cho, S., Crenshaw, K. W., & McCall, L. (2013). Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs: Journal of Women in Culture and Society*, *38*(4), 785–810. https://doi.org/10.1086/669608

Colombo, E., & Rebughini, P. (2016). Intersectionality and beyond. *Rassegna Italiana Di Sociologia*, *3/2016*. https://doi.org/10.1423/84373

Crenshaw, K. W. (1994). Mapping the margins: Intersectionality, identity politics, and violence against women of color. In *The Public Nature of Private Violence*. Routledge. ISBN: 978-0-203-06090-2

Davis, A. Y. (1983). *Women, race & class*. Vintage. ISBN: 978-0-307-79849-7

Djoudi, H., Locatelli, B., Vaast, C., Asher, K., Brockhaus, M., & Basnett Sijapati, B. (2016). Beyond dichotomies: Gender and intersecting inequalities in climate change studies. *Ambio*, *45*(3), 248–262. https://doi.org/10.1007/s13280-016-0825-2

Eurostat. (2003). *Household budget surveys in the EU — Methodology and recommendations for harmonisation – 2003*. ISBN: 92-894-5435-0

Goodman, A., & Oldfield, Z. (2004). *Permanent differences? Income and expenditure inequality in the 1990s and 2000s* [Research Report]. *R66*. https://doi.org/10.1920/re.ifs.2004.0066

Kaijser, A., & Kronsell, A. (2014). Climate change through the lens of intersectionality. *Environmental Politics*, *23*(3), 417–433. https://doi.org/10.1080/09644016.2013.835203

Tomás, M., García-Muros, X., Alonso-Epelde, E., Arto, I., Rodríguez-Zúñiga, A., Monge, C., & González-Eguino, M. (2023). Ensuring a just energy transition: A distributional analysis of diesel tax reform in Spain with stakeholder engagement. *Energy Policy*, *177*, 113558. https://doi.org/10.1016/j.enpol.2023.113558

Van Beek, L., Hajer, M., Pelzer, P., Vuuren, D. van, & Cassen, C. (2020). Anticipating futures through models: The rise of integrated assessment modelling in the climate science-policy interface since 1970. *Global Environmental Change*, *65*, 102191. https://doi.org/10.1016/j.gloenvcha.2020.102191

Walker, G. (2010). Environmental justice, impact assessment and the politics of knowledge: The implications of assessing the social distribution of environmental outcomes. *Environmental Impact Assessment Review*, *30*(5), 312–318. https://doi.org/10.1016/j.eiar.2010.04.005

Weyant, J. (2017). Some contributions of integrated assessment models of global climate change. *Review of Environmental Economics and Policy*. https://doi.org/10.1093/reep/rew018