Applications and Examples

Gherardo Varando

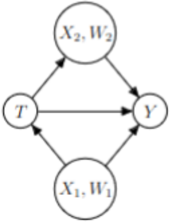
IPL

16 May 2023

Common Driver example

Kretschmer et al. [2021]

Double Machine Learning



$$Y = \theta(X) \cdot T + g(X, W). \tag{1}$$

The parameter θ describes the direct effect of some variable T on the outcome variable Y.

$$Y = \theta(X) \cdot T + g(X, W) + \epsilon \qquad \mathbb{E}[\epsilon | X, W] = 0 \tag{2}$$

$$T = m(X, W) + \eta \qquad \mathbb{E}[\eta | X, W] = 0 \tag{3}$$

$$\mathbb{E}[\eta \cdot \epsilon | X, W] = 0. \tag{4}$$

We proceed according to the partialling out method of the double machine learning framework Chernozhukov et al. [2018]:

1. Fit an estimator $\mathbb{E}[Y|X, W]$ of Y on X and W,
2. fit an estimator $\mathbb{E}[T|X, W]$ of T on X and W,
3. compute residuals over $\tilde{Y} = Y - \mathbb{E}[Y|X, W]$ and $\tilde{T} = T - \mathbb{E}[T|X, W]$ and
4. estimate $\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_i (\tilde{Y} - \theta(X) \cdot \tilde{T})^2$.

The Q10 model

$$R_{eco}(X, T) = R_b(X, T) \cdot Q_{10}^{(T - T_{ref})/10}, \tag{5}$$

Following the example of Reichstein et al. [2022], we used data from the EC tower in Neustift, Austria, from 2003 to 2007 of the FLUXNET2015 dataset and generated the data similarly.

Synthetic data

$$RECOsyn = Rbsyn \cdot Q_{10}^{0.1 \cdot (T - 15)}, \tag{6}$$

$$Rbsyn = 0.75 \cdot (\tilde{R}_b^{syn} - \min(\tilde{R}_b^{syn}) + 0.1 \cdot \pi), \tag{7}$$

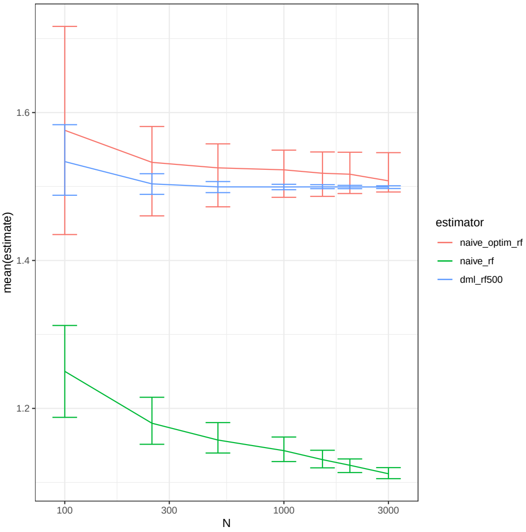
$$\tilde{R}_b^{syn} = 0.01 \cdot SWPOTsm - 0.005 \cdot SWPOTsmdiff, \tag{8}$$

- Rb^{syn} describes the base respiration
- The smooth incoming potential radiation $SWPOTsm$ and its smoothed difference quotient $SWPOTsmdiff$ are computed by averaging moving windows of 10 days over the incoming potential radiation SW_{POT}.
- The Q_{10} temperature coefficient is set to 1.5.

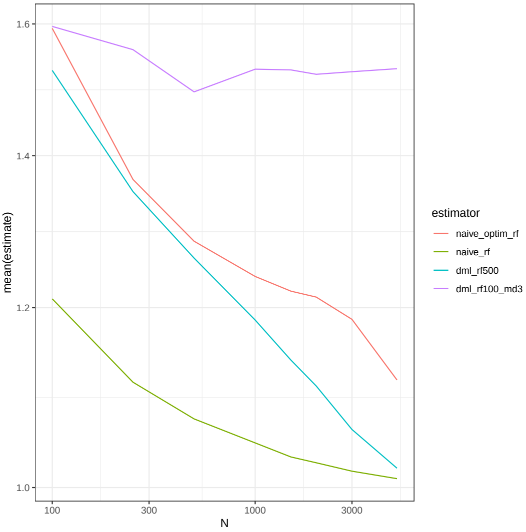
In-situ data

Ecosystem respiration is not directly observed at flux towers during the day. It is a latent flux that can only be measured under controlled conditions such as a sealed chamber. At night, however, we assume GPP to be zero as there is no photosynthesis occurring, and all carbon flow stems from respiration. From all measured data, the ones that do not fulfill a certain quality criterion are filtered out and gap filled. We will work with measured data only. Only about 10% of the nighttime data is observed. Thus, we estimate Q_{10} on 4331 data points. As predictors we will use the day of the year to account for seasonality, vapor pressure deficit VPD, and soil water content SWC.

Bootstrapped results for synthetic data

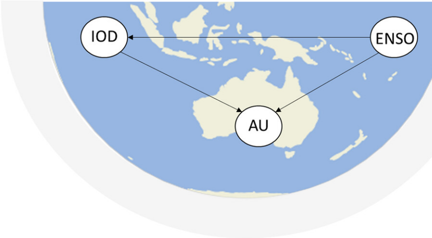


Bootstrapped results for real data



ENSO Effects on Spring Precipitation in Australia

- Effect of El Niño Southern Oscillation (ENSO) on Australian precipitation (AU) during spring, and the possible mediation of the Indian Ocean Dipole (IOD) [Kretschmer et al., 2021]
- monthly NCEP reanalyses covering 1949–2019
- ENSO (Niño, neutral, Niña), IOD into positive (+), neutral (0) and negative (-) phases; and AU into above (high) and below (low) average values



Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018. ISSN 1368-4221. doi: 10.1111/ectj.12097. URL https://doi.org/10.1111/ectj.12097.

M. Kretschmer, S. V. Adams, A. Arribas, R. Prudden, N. Robinson, E. Saggioro, and T. G. Shepherd. Quantifying causal pathways of teleconnections. *Bulletin of the American Meteorological Society*, 102(12):"E2247–E2263, 2021.

M. Reichstein, B. Ahrens, B. Kraft, G. Camps-Valls, N. Carvalhais, F. Gans, P. Gentine, and A. Winkler. Combining system modeling and machine learning into hybrid ecosystem modeling. In *Knowledge-Guided Machine Learning*. 2022. doi: https://doi.org/10.1201/9781003143376-14.