WAGENINGEN
UNIVERSITY & RESEARCH

Transformative ways to study transformative bioeconomies

Final wildcard project report

Marcel Kornelis, Xuezhen Guo, Koos Gardebroek, Charlotte Harbers, and Daoud Urdu

31-12-2023

TRANSFORMATIVE Bioeconomies

1. Transformative ways to study transformative bioeconomies

Innovative idea and objective

Econometric models can be used to asess the impact of the various driving forces of bioeconomies and/or predict the future of bioeconomic indicators. In these settings, it is common practice to only include driving forces as predictors in the model if statistical tests indicate significant relationships between these forces and the performances of the bioeconomic variables of interest. It is a major challenge to statistically reveal significant relationships if datasets are limited in size, especially when complex non-linear patterns or relationships exist among the predictor variables. Recently developed decision-tree based machine-learning techniques such as random forests can handle a large number of used predictors, can capture various types of interrelationships, and can deal with abrupt regime shifts, even when the available datasets are limited in size.
Our objective is to predict availability of raw materials for biobased products, and our idea is to investigate if we can extend the traditional model building process with machine-learning techniques, to see if this improves the predictive power of the econometric models if the datasets are limited and complex non-linear pattern exists.

Relevance to the materials transition in textiles and/or building materials?

Within the context of transformative bioeconomies, one may expect a fundamental changing world in the near future for at least two reasons. First, if interventions by decision makers, that help to phase out of fossil feedstock, occur to be successful, it may lead to a fundamental change in the structure of the bioeconomic system. Second, very recently occurring historic events at the geopolitical level may lead to worldwide regime shifts that may also alter the future performance of bioeconomies. Since, our current position is at the very beginning of such a regime shift, actual data that contain information about this transformative process can be (very) limited. This implies that it becomes relevant to make optimal use of the available data if one wants to understand and predict the transformative process. This notion underpins our methodological approach.

What did you do?

We investigated the use of mixed effects methods for panel (longitudinal) data complemented with a random forest algorithm (MERF) to model bioeconomies. In a workshop for bioeconomy experts in the textiles and building sectors, we proposed, explained, and discussed our methodology. Two empirical cases were selected: predicting the acreage use of flax and hemp together in European countries, and predicting cotton production in Asian countries. In the analysis, we considered various types of MERF models, that ranged from very parsimonious models without a random forest component to more complex models, in which the random forest algorithm was an integral part of the model building process. For each of the developed models, we checked its predictive power and compared it with the other model alternatives.

Main result, achievement and highlight

We found the following results:

- The combination of methods provided better predictions than the traditional model without the random forest algorithm.
- The random forest algorithm including additional variables that were found to be insignificant in the traditional model can improve the prediction. This implies that the random forecast model is good at capturing the complex non-linear patterns in the data, which were hard to be captured by the traditional model.
- The random forest algorithm remains a black box, which makes this component difficult to interpret.
- The model could well capture the heterogeneity among countries, which again shows the advantage of complex relationship capturing for the random forest algorithms

Key message

In this project, we learned that in small datasets, traditional models often indicate that interrelations among variables are insignificant. However, including these variables to a random forest component that is added to the model gives extra predictive power. So, there seems to be relevant relationships in the data which are not captured with traditional estimation techniques (which are essentially additive linear modelling techniques)

We would like to generalize our findings in the following ways:

- Apply the methodology to other datasets to investigate the general applicability of MERF.
- Change the convergence levels of the random forest algorithm to better understand its robustness.
- Trying to make the random-forest blackbox more interpretable.

Visual abstract

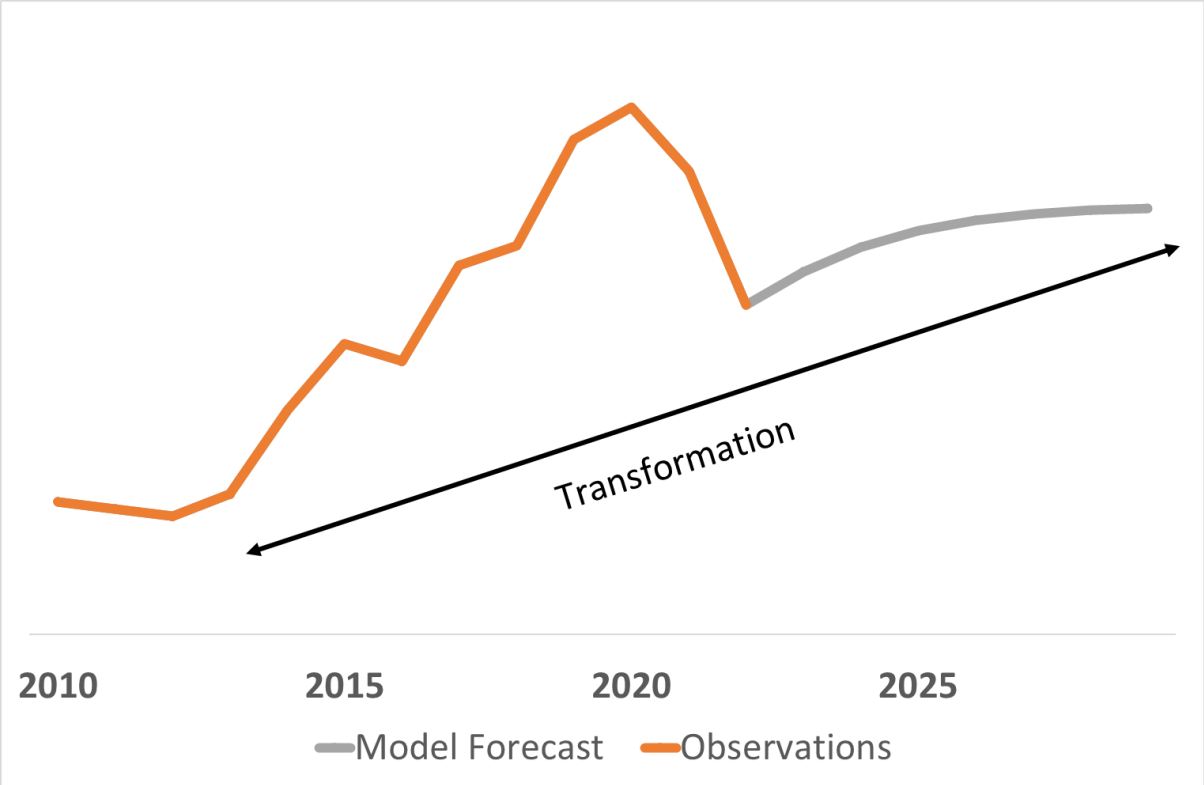

Figure. Illustration of a transformative market

2. Additional questions about progress and 'readiness'

Where you started

The methodology of the traditional mixed effects model follows already existing procedures. This is also the case for the mixed effects model with a random forest algorithm (MERF), as we found a number of references in other disciplines that follow this practice. We used these references as a starting point to program the whole model building processes in Python.

Where are you now

We developed the MERF model building process for the prediction of acreage use of hemp and flax together. This was our main goal, and we managed to develop it. Our intention was to also develop such a model building process for the prediction of the acreage use of cotton among Asian countries. This appeared more difficult than expected, and we are still working on this.

Potential and next steps

The potential of our project to contribute to the materials transition is that we are developing a methodology that can make optimal use of newly available information to predict or simulate the over-time effectiveness of unexpected events or planned interventions regarding bioeconomies. We think that a logical next step is to have a better understanding

of what the machine learning techniques is doing with the available information, so that we better understand why their use is an improvement.

Innovation readiness

We believe that the technology readiness is at the technology demonstration level (6), and the societal readiness is at the level of co-operation with relevant stakeholders (5).

3. Learning Journey

1. Did your Wildcard project involve new collaboration with disciplines or people? If so, briefly explain what was new.

Yes, between AFSG (Xuezhen Guo & Charlotte Harbers) and DMW (Koos Gardebroek) , and between AFSG(Charlotte Harbers) and WEcR (Marcel Kornelis). New was using the same methods in different application environments.

2. If applicable, did the new collaboration alter your original thinking about the topic? Did it change research directions or courses of action? If so, briefly characterize how.

No, it did not. We found out that although different disciplines may use different jargon, in the end, if we considered the courses of action from a higher abstract level, the original thinking of the different disciplines were very similar.

3. Did interactions during community days and/or meetings organized by the investment theme alter your original thinking about the topic? Did such interactions change research directions or courses of action? If so, briefly characterize how.

   Yes, we shared ideas about how to model diffusion of innovations, the role of economics therein, and which models to consider. Very fruitful exchanges.

4. Did you meet any challenges during implementation of your wildcard project? If so, what kind of challenges where these?

   The available data turned out to be more limited than expected (fewer observations than hoped for). This did not prevent us from analyzing the models envisioned. Instead, we made it our main challenge! It was also challenging to capture all model complexities.

5. If applicable, how were these challenges eventually addressed? Did activities organized by the investment theme contribute to overcoming challenges? If so, briefly indicate how.

   By good internal discussions, we came up with a step-wise approach to deal with this. In one of the last meetings, we also agreed to apply the methodology on an existing more extensive dataset for biofuel crops.

6. Has your involvement in the investment theme resulted in any new initiatives or spin-offs that would probably not have emerged if you had not participated? If so, briefly indicate how these new initiatives came about.

Yes, certainly. Without this project we would not consider comparing traditional panel data models with machine learning versions. Our learning experiences have already been used in a bilateral project from WFBR about applying machine learning time series approach to predict fruit prices; in an investment theme project in which we apply spatial-temporal machine learning techniques to forecast disease spread over time globally; and in an investment project about nonlinear diffusion of innovative markets. Finally, the outcomes are planned to be used in a PhD course. We also got an idea for a follow-up scientific paper.

4. Additional project specific deliverables

Copy-paste the deliverables provided in your submission document and explain how you have met these deliverables. If deliverables could not be reached, please explain.

Additional deliverables proposed when submitting the Wildcard project

	Activity	Deadline
1	Developing a methodology that can be used to determine to what extent a specific approach (e.g. expert opinions, parametric models, etc.) is fast and flexible enough to adopt new information.	May 2023
2	Workshop for bioeconomy experts in the textiles and buildingsectors to explain the developed methodology, to receive feedback to improve it, and to select an empirical case	Beginning of June 2023
3	Upgrade the methodology based upon the workshop	August 2023
4	Apply the methodology to the selected empirical case of the workshop	December 2023
5	Write a manuscript about the results of the empirical case (we write the manuscript parallel to the empirical analysis)	December 2023

Status of each project specific deliverable

	Activity	Status
1	Developing a methodology that can be used to determine to what extent a specific approach (e.g. expert opinions, parametric models, etc.) is fast and flexible enough to adopt new information.	Done
2	Workshop for bioeconomy experts in the textiles and buildingsectors to explain the developed methodology, to receive feedback to improve it, and to select an empirical case	Done
3	Upgrade the methodology based upon the workshop	Done, we programmed the methodology in STATA and Python
4	Apply the methodology to the selected empirical case of the workshop	In the workshop, it was decided to first focus on hemp and flax, and after that on cotton. The hemp-and-flax case is finished, the cotton case has a delay, we informed Katharina and Cees about this delay
5	Write a manuscript about the results of the empirical case (we write the manuscript parallel to the empirical analysis)	We'll send this in, as soon as we added the cotton case to the manuscript