



Exploring Possible Futures

Video Transcript

Data handling

When it comes to designing your own models, you will likely stumble upon data issues sooner rather than later. In this lecture, we will have a short look at some points you want to consider when handling your model's data.

With conceptual models, the mathematics and the interpretation are the main output you're aiming for. Even if often small numerical examples are used for illustration, they are not that demanding in terms of data management. But with numerical modelling, you're looking for quantifications. And the solution of your model will consequently provide numbers.

Lots of numbers.

And to get those, you will have to put in numbers first. Let's take our simple Optimus system model as an example. Before getting any model outcomes, you will have to define the different cost parameters, the c_{var} and the c_{invest} , the price for emissions, the t , and emission factors, the em_k , for each of the technologies k you want to consider, the time structure I of your model and the resulting demand levels, the dem . Once you have all those and your model can finally run, you will get the values for the generation, the q , and the installed capacities, the q_{max} . This may sound rather unproblematic.

But if you design an energy system model on an hourly basis for investment decisions up to 2050, those numbers add to huge data sets in no time. What data do we typically need to design a running numerical model? As those models are normally representations of markets or systems, we basically need the structural elements describing our system. This can include a given market infrastructure, like production or transport facilities, as well as physical aspects like input-output relations, construction and lifetimes, and cost structures. Similar, we may need definitions of the respective environment we have included in our model, like emission factors or the assimilative capacity of an ecosystem.

Contrary to those structural data, we have the operational data those systems produce-- namely output-- and prices. Often, our model will have those elements as endogenous variables and we don't need them as direct input. But it's always helpful if we have a real-world observation we can calibrate our model to. Similar, if the model is focused on one single market, the output and prices of another market may be a needed input for our model. For example, coal, oil, and gas prices are typically needed for electricity market modelling. Where can you find the data you need? There are plenty of sources out there: statistical offices of the respective countries, international organisations, yearly company reports, and so forth.

Gathering the data is often a bit frustrating. And sometimes, you simply cannot find the data needed. But don't give up yet.

To fill the data gaps of your model, you may be able to extrapolate the data you have or derive meaningful assumptions by using few data points and economic theory. For example, future investment costs can be constructed by assuming cost-reduction levels based on historic cost reductions. Or, you can construct a demand function by a few observations of real-world demand and price pairs.



The most important aspect of those derivations and assumptions is to make clear what you do and what your assumptions are. When it comes to your model results, there are two important rules you should take care of before making the effort of writing up your interpretation. Rule number one relates to the fact that we design our own models, and thereby should have a good idea in which direction the results are likely to develop. Thus, if you are really surprised by your findings, chances are high that they are caused by some model error. Maybe there's a problem with the data upload or the model code and the model you think you've coded are not the same.

So take really care to figure out what drives your results. Those errors can be hidden really well. Rule number two is basically the same story. Even if your results look reasonable, there's always a chance that there's a model error in the background. I have already written up whole results sections multiple times, every time with really good-looking results. But there was a data upload error once and a misplaced bracket the second time. Well, that's the life of a modeller. So, essentially, make sure you know your model is working as you think it should work. That's why you should always make a small, traceable model example first and then go big later.

If you're convinced enough that your results are correct, you'll likely end up with up to millions of data points. Remember, we want to connect those with our problem. So we need to interpret our results with our problem in mind. To do so, we need number crunching. On the one hand, we need to filter all our results and boil them down to meaningful aggregates, or representative cases, maybe deriving yearly average prices out of hourly prices or aggregate locational information to regional or national totals. On the other hand, we need to present them. Numerical models have the advantage that they are well-suited to produce simple, yet meaningful, graphical representation of complex processes.

Remember, a single picture can be worth a thousand data points.