

# Full lifecycle of a nowcasting project

## UNCTAD's work and experience

Daniel Hopp

UNCTAD

25 May 2022



- 1 Is nowcasting applicable
- 2 Performing the analysis
- 3 Automating the nowcasts
- 4 Final product
- 5 Conclusion and resources

- 1 Is nowcasting applicable
- 2 Performing the analysis
- 3 Automating the nowcasts
- 4 Final product
- 5 Conclusion and resources

## Characteristics of a nowcasting analysis

- Relevant if you have an existing indicator that is published with a significant lag and want timelier estimates
- Two additional key characteristics for your target variable:
  - Must have a sufficient number of observations to effectively train a model/establish relationship to input variables
  - Must exist explanatory variables that are both related to the target variable and published in a more timely manner

## Example of SDG indicators

- Initial feasibility analysis explained much more thoroughly in a recent [UNCTAD research paper](#)
- In addition to laying out the conditions for a successful nowcasting exercise, it carries out a comprehensive feasibility survey for all SDG indicators and sub-indicators
- Documents their publication history, publication lag, and the potential existence of explanatory variables

- 1 Is nowcasting applicable
- 2 Performing the analysis
- 3 Automating the nowcasts
- 4 Final product
- 5 Conclusion and resources

## Step 1: identifying and gathering the data

- Write a list of any variables that could possibly help predict the target variable, the more the better
- From the list, research specific variables/sources, eliminating based on publication history, publication lag, ease of extraction
- Download/extract the data for all the candidate explanatory variables
- Not essential, but good at this point to already consider automation/reproducibility in the extraction process
- Focus first on large, standardized data sources first (OECD, World Bank, etc.), where a large quantity of variables is available with minimal marginal effort
- Be wary of spending much time on bespoke / specific data sources, where you may waste time extracting the data for a single series that may or may not help the model in the end

## Step 2: transforming and organizing the data

- Ideally integrate this step with the previous, gathering all series into one dataframe with observations in rows and different series in columns
- A key or catalog file can help with this, dictating series names, sources, periodicity, required transformations, etc.
- Apply necessary transformations to series, main ones are usually removing seasonality and converting to growth rates/stationarity
- The necessity of the stationarity transformation depends on the chosen methodology, but is a requirement for many and does not hurt for those that don't require it, so safest to perform this conversion

# Example dataset

date	mar_freight	mar_freight	tourists_de	tourists_es	tourists_fr	tourists_it
01.01.02	NA	NA	NA	NA	NA	NA
01.02.02	NA	NA	0.01267887	0.00903954	NA	0.00319619
01.03.02	NA	NA	-0.0026893	-0.0079574	NA	-0.0034945
01.04.02	NA	NA	0.02310277	0.00355651	NA	0.01896985
01.05.02	NA	NA	-0.0187558	-0.0055314	NA	9.21E-05
01.06.02	0.01816678	0.03723896	0.01384783	-0.0110961	NA	-0.0251099
01.07.02	NA	NA	-0.0023298	0.00771741	NA	0.03175989
01.08.02	NA	NA	0.00783221	-0.0039522	NA	-0.0102255
01.09.02	0.02089536	-0.0282957	-0.0054091	0.00978129	NA	0.00951461
01.10.02	NA	NA	-0.0121052	0.00313769	NA	-0.013169
01.11.02	NA	NA	0.01586572	-0.0016316	NA	1.49E-05
01.12.02	0.03317017	0.04097487	-0.0052287	-0.0021612	NA	-0.0129945
01.01.03	NA	NA	0.02173766	0.01264906	NA	0.00555717
01.02.03	NA	NA	-0.0251749	-0.0159919	NA	-0.0084934
01.03.03	-3.67E-04	-0.0562409	0.01140295	0.00405357	NA	-0.0024096
01.04.03	NA	NA	-0.0178507	0.01217669	NA	-0.0073873
01.05.03	NA	NA	0.012514	6.17E-04	NA	-0.2055454
01.06.03	-0.0053868	0.05011756	0.01628325	0.02002472	0.05407676	0.30028286
01.07.03	NA	NA	0.00171155	-6.38E-04	-0.0209694	-0.0457787
01.08.03	NA	NA	0.01004855	-0.0091055	0.00962922	0.02084191
01.09.03	-0.0021222	-0.0351053	2.42E-04	-0.0281883	-0.0182409	-0.0285667

## Step 3: selecting a methodology/methodologies

- Look at this [UNCTAD research paper](#) as a starting point if you don't know which methodology to use
- Paper benchmarks most commonly used nowcasting methodologies on US GDP growth and includes example code which you can tailor to your needs
- Some commonly used approaches:
  - dynamic factor models (DFM)
  - long short-term memory artificial neural networks (LSTM)
  - mixed-data sampling regression (MIDAS)
  - Bayesian vector autoregression (Bayesian VAR)

## Step 4: performing the modelling exercise

- With your chosen methodology, set up a training and testing loop
- Train with a certain percentage of your data, assess its performance on a test set (e.g., with RMSE)
- Setting up plotting to visualize predictions on the test set also helps
- If your methodology has hyperparameters, a validation set may also be necessary
- Helps to set up a standardized results output for each model run, detailing the algorithm used, the variables in the model, the model's hyperparameters, and performance
- Idea is to be able to recreate any model based on the information in the results file and compare permutations' performance to select best one

## Step 4: performing the modelling exercise (cont.)

- Another important aspect in nowcasting is assessing on synthetic data vintages
- Your model may perform very well when it has all data available to it for the time period
- In practice, this situation only occurs near the end of a prediction period
- To ensure robustness even early in a prediction period, when little information is available, must test artificially introducing missing data
- The *nowcast\_lstm* library has a provision for automatically selecting the variables that go into your model as well as hyperparameters
- For other methodologies, you can use techniques like recursive feature addition (RFA), recursive feature elimination (RFE), random sampling, or others

## Step 5: finalizing the model

- Previous step will take much time, trial and error, exploration
- A final model just means settling on:
  - A methodology
  - Its hyperparameters (if any)
  - The variables that go into the model

- 1 Is nowcasting applicable
- 2 Performing the analysis
- 3 Automating the nowcasts**
- 4 Final product
- 5 Conclusion and resources

## Automation: general comments

- A nowcast has to be updated on a regular basis
- Avoid anything manual anywhere in your process, will quickly become burdensome
- Avoid any absolute references anywhere in your process, e.g., "C:/Users/Me/Projects...", ideally should be a repo that can be cloned and run anywhere, or even better containerized with Docker
- Avoid hardcoding anything, make ample use of key files to store things like required transformations, hyperparameters, target variables, etc.
- Host your flow on a server and schedule it with Airflow or a cron job with error logging

## Automation: general comments (cont.)

- Modularize the flow, helps with maintenance, three main components:
  - data gathering and transformation, i.e., output an aggregated latest data file or populate a database
  - running the model, i.e., taking that data file and retraining the model and generating latest predictions in a file
  - final UI output, i.e., taking the prediction file and creating a document, dashboard, website, etc. from it

- 1 Is nowcasting applicable
- 2 Performing the analysis
- 3 Automating the nowcasts
- 4 Final product**
- 5 Conclusion and resources

## No dissemination

- Don't plan to use the nowcasts as a standalone output, just need to have as inputs for internal work or for use when necessary in other reports
- No further work required after automation, access to the latest nowcasts is sufficient

## Periodic report

- Publishing a periodic report with some analysis on the latest nowcasts
- UNCTAD's current approach with [nowcast bulletins](#)
- Allows in-depth commentary on the nowcasts at a given point in time
- Drawback of presenting a dynamic, developing number at one single (arbitrary) point in time, can misconstrue the nature of the nowcast and falsely lead to belief that the figure is the definitive prediction of the organization

## Dynamic, interactive website

- Website/dashboard that is updated in real time
- Always reflects the latest nowcast
- Can add interactive elements such as tooltips for components' contributions to updates in the nowcast
- More in line with the nature of nowcasts than a static, periodic page
- UNCTAD developing a [website](#) to be launched in June 2022, modelled after [FED's nowcast web site](#)

- 1 Is nowcasting applicable
- 2 Performing the analysis
- 3 Automating the nowcasts
- 4 Final product
- 5 Conclusion and resources

## Conclusion

- Many components to creating a successful nowcasting application
- Resources discussed should significantly help with the exploration and development phase
- Keep in mind the automation and coding best practices mentioned in the presentation
- Bigger upfront investment, more planning required, but greatly contributes to sustainability, extendibility, and robustness of the application

## Resources

- [SDG survey UNCTAD research paper](#)
- [Benchmark UNCTAD research paper](#)
- [UNCTAD nowcast bulletins](#)
- [Benchmark repository](#)
- [FED nowcasting website](#)
- [UNCTAD nowcasting website \(WIP\)](#)