Capstone Project Summary

Choosing the Best Model – An Out-of-sample Testing for Gradual Physical Impact Model

By

Xinqi Wang

Submitted to
Central European University
Department of Economics and Business

In Partial Fulfilment of the Requirements for the degree of Master of Science in Business Analytics

Budapest, Hungary
2021

Table of Contents

Table of Contents ii
Background 1
Problem 1
Methodology and Key Outcomes 2
Learning Experience and Future Improvements 3

Background

Awareness of climate risk is rapidly changing, and it is also reshaping the global economy. However, there is a considerable level of uncertainty in understanding the macroeconomic consequences of climate change. According to the Swiss Re Institute[1], the global temperature could rise by more than 3°C and cause the world's GDP to decrease by 18% if no mitigating action is taken by 2050. On the other hand, if the Paris Agreement targets are met, the impact can be lessened at 4% decrease in the predicted global GDP by 2050. Numerous past studies focus on understanding the relationship between gradual physical impacts and economic growth.

My client Ortec Finance is a global provider of technology and solutions for asset-liability management, risk management, financial planning and portfolio constructions. They design, build and deliver high-quality software models to provide their clients with new ways to solve complex investment decision-making difficulties. I work with their climate ESG solutions team in my capstone project to help them with empirical work on an existing gradual physical impact model.

For modelling gradual physical impacts (the impact of gradual temperature increases and changes in precipitation on economic output per region), they draw from the work of Burke and Tanutama (2019)[2], which builds on earlier work by Burke et al. (2015[3], 2018[4]). Burke et al. (2015, 2018) use country-level data on economic aggregates (e.g., gross domestic product) to study total damages. Burke and Tanutama (2019) use a district-level panel dataset on climate and GDP across 37 countries and multiple decades, and confirmed that also on a district level, economic production is concave in temperature exposure, with a negative slope throughout nearly all the observed temperature distribution and increasingly steep at warmer temperatures.

Problem

To date, my clients have based their approach on the baseline model without temperature lags from Burke and Tanutama (2019). However, they wonder if it might be more appropriate to apply the 1or 5 lag model; in this way, longer-term temperature effects can be captured. For lag modelling, the data must cover a sufficiently long period to determine lag effects. Burke and Tanutama (2019) use 154 thousand district-year observations from 11,000 districts. This means that they cover 14 years of data, which is not enough to build a sound 5-year lag model. So, from an econometric perspective,

[1] The Swiss Re Institute. (2021). World economy set to lose up to 18% GDP from climate change if no action taken, reveals Swiss Re Institute's stress-test analysis.

[2] Burke, M., and Tanutama, V. (2019). Climatic Constraints on Aggregate Economic Output. No. w25779. National Bureau of Economic Research.

[3] Burke, M., Hsiang, S.M. & Miguel, E. (2015). Global non-linear effect of temperature on economic production. Nature 527, 235-239.

[4] Burke, M., Davis, W.M. and Diffenbaugh, N.S. (2018). Large potential reduction in economic damages under UN mitigation targets. Nature 557.7706 (2018): 549.

this rules out the possibility of using lag 5 model on this dataset. However, a one-year lag could be appropriate. To determine whether the lag 1 model is an appropriate model to use, I would ideally perform out-of-sample testing. As a result, the objective of my capstone project is to find out whether a lag model (1 or more lags) is more appropriate to estimate future gradual physical impacts on GDP than the baseline model without temperature lags. The desired outcome would result from out-of-sample testing and arguments for either switching to a lag model or sticking with the no lag model.

Methodology and Key Outcomes

Before diving into model analysis, I conduct a long list of relevant literature to catch up with the latest methodology and findings. For example, Kahn et al. (2019)[5] use an econometric model analysis approach to quantify the effects of climate change on economic growth. They use a stochastic growth model. A panel data set of 174 countries from 1960 to 2014 and find an average of 0.04°C increase in temperature per year will reduce income growth by 7.22% per year by 2100 there are no mitigation policies involved. On the other hand, if the temperature increase can be limited to 0.01°C per year by compiling the Paris Agreement, the reduction of income growth will be reduced to 1.07%.

In my study, I use annual climate data and real GDP per capita and follow the methodology from Burke et al. (2018). I choose the sample period from 1961 to 2019 and build an unbalanced panel dataset with 37 countries because of the lack of GDP data. The explained variable is the growth rate of log GDP per capita. It is computed using the country's annual purchasing power parity-adjusted GDP and population, calculating the first difference and then taking logs. For the independent variables, I use both temperature and its square, and the same with precipitation. They are aggregated from the monthly average temporally to the annual level and spatially at the country level. The unit of temperature data in Celsius and precipitation is in millimetre (mm). These data are then merged with GDP growth data. The source of the GDP data is The DataBank. Temperature and precipitation data are from The World Bank Group Climate Change Knowledge Portal.

After collecting all the needed data, I first perform an in-sample analysis. Next, I estimate a panel fixed effects model developed by Burke et al. (2018) to understand the historical relationship between temperature and economic output. When the temperature is below 65.5°C, holding everything else constant, the increasing temperature positively affects GDP growth but at a diminishing rate. On the other hand, when the temperature is above 65.5°C, holding everything else constant, the increasing temperature has a negative effect but at an increasing rate, meaning the damage would grow larger and larger as time passes. This result is also statistically significant. Next, I perform rolling out-of-sample forecasting to determine whether the lag 1 model is an appropriate model to use and whether country-specific linear and quadratic time trends are necessary. Finally, I

[5] Kahn et al. (2019). Long-Term Macroeconomic Effects of Climate Change: A Cross-Country Analysis. NBER

choose a continuous window size from 10 to 30 and find the dynamic panel model performs the best among all the models. To test whether the outperformance of the dynamic panel model is statistically significant, I choose to perform a Giacomini and White test. The test results confirm that this outperformance is statistically significant at 1%, 5% and 10% when comparing with different models using different window sizes.

Given the empirical results, I conclude that the dynamic panel model has the most downward bias and the best forecasting power than the static panel model. Therefore, I advise my client to switch to a dynamic model, especially a lag 1model from the previous static model.

Learning Experience and Future Improvements

This project has been a great learning experience especially for applying the data analysis knowledge I learned from this program. Coding-wise, I used Stata to perform the in-sample analysis, MATLAB for the out-of-sample forecasting, and R for data cleaning, waggling, and visualization. Courses I took for using R in data analysis have been beneficial along my capstone journey. Packages like tidyr, dplyr, and stringr played an essential part in my data cleaning and waggling phases.

Due to the time limit, I could not perform a population-weighted climate data analysis done by Kahn et al. (2019). Scaled climate data will not have a larger country bias. One would naturally think that larger countries have much more human damages to the environment than those with a smaller population and, therefore, a higher growth rate in temperature raising. As a result, constructing population-weighted climate data would solve this problem and provide a more robust result. Further research is encouraged on population-weighted climate data and extended time series data and different countries as I leave the findings open to interpretation.